

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: July 20, 2005, 18:16:40 ; Search time 118 seconds

(without alignments)  
1358.312 Million cell updates/sec

Title: US-10-017-471b-17

Perfect score: 1645  
Sequence: 1 MPEAVVLINSASDANSIEQT.....GHODGLVPLTTLGPAFSG 313

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot\_03.\*

1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1634.5	99.4	314	2	086851 streptomyc
2	1023.5	62.2	301	2	AFSA STRGR
3	602	36.6	309	2	056152 streptomyc
4	562	34.2	294	2	024738 streptomyc
5	524.5	31.9	331	2	083X19 streptomyc
6	510	31.0	317	2	084H95 streptomyc
7	480.5	29.2	291	2	024740 streptomyc
8	319.5	19.4	345	2	082KU4 streptomyc
9	271.5	16.5	353	2	08JN90 streptomyc
10	132	8.0	1199	2	Q7NCZ7 streptomyc
11	120.5	7.3	597	2	Q727P2 streptomyc
12	118.5	7.2	563	2	Q8N510 streptomyc
13	118.5	7.2	563	2	Q8N510 streptomyc
14	109.5	6.7	1762	2	Y084 HUMAN
15	106	6.4	372	2	06Z2G9 streptomyc
16	105.5	6.4	942	2	Q9L114 streptomyc
17	104.5	6.4	607	2	Q82M25 streptomyc
18	103.5	6.3	671	2	Q925X4 streptomyc
19	103.5	6.3	1089	2	Q7X2F4 streptomyc
20	101.5	6.2	1002	1	NM3B RAT
21	101	6.1	247	2	Q6AKF3 streptomyc
22	101	6.1	500	2	Q89LE3 streptomyc
23	101	6.1	548	2	Q64QC7 streptomyc
24	101	6.1	6396	2	Q9KID7 streptomyc
25	100.5	6.1	292	2	Q82HN2 streptomyc
26	100.5	6.1	777	2	0826D8 streptomyc
27	100	6.1	1292	2	Q758Q9 streptomyc
28	99.5	6.0	802	2	Q8KJH4 streptomyc
29	99	6.0	425	2	Q84W44 streptomyc
30	99	6.0	549	2	Q82BEO streptomyc
31	99	6.0	4391	1	PGEM_HUMAN

32	98.5	6.0	193	2	Q69PT0 streptomyc
33	98.5	6.0	2747	2	Q9L800 streptomyc
34	98	6.0	259	2	Q751L4 streptomyc
35	98	6.0	1408	2	Q6X125 streptomyc
36	97.5	5.9	274	2	Q62D39 streptomyc
37	97.5	5.9	274	2	Q63M77 streptomyc
38	97.5	5.9	4557	2	Q8CJN6 streptomyc
39	97	5.9	1480	2	Q8GC13 streptomyc
40	96.5	5.9	263	2	Q9ADT4 streptomyc
41	96.5	5.9	495	1	A1BG HUMAN
42	96.5	5.9	5060	2	Q52545 streptomyc
43	96	5.8	504	2	Q82DTS streptomyc
44	96	5.8	725	2	Q7SEH9 streptomyc
45	96	5.8	1802	2	Q6V1N7 streptomyc

## ALIGNMENTS

RESULT 1	ID	PRELIMINARY	PRT	314 AA
AC	086851			
DT	01-NOV-1998 (TREMBLrel. 08, Created)			
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)			
DT	25-OCT-2004 (TREMBLrel. 28, Last annotation update)			
DE	ScdA protein.			
GN	Name=ScdA; ORFNames=SCAH10.31;			
OS	Streptomyces coelicolor.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Streptomycinae; Streptomycetaceae; Streptomyces.			
NC	NCBI TaxID=1902;			
LN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Subacteria, and M145;			
RA	Takano E., Chakraborty R., Nihira T., Yamada Y., Bibb M.,			
RL	Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.			
LN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=A3(2) / M145;			
RX	MEHLIN=2196410; PubMed=12000953; DOI=10.1038/417141a;			
RA	Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,			
RA	Thomson N.R., James K.D., Harris D.B., Quail M.A., Kleser H.,			
RA	Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,			
RA	Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,			
RA	Huang C.-H., Kleser T., Larke L., Murphy L.D., Oliver K., O'Neill S.,			
RA	Rabinowitz E., Rajadream M.A., Rutherford K.M., Rutter S.,			
RA	Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,			
RA	Warren T., Wietzorek A., Woodward J.R., Barrell B.G., Parkhill J.,			
RT	Hopwood D.A.;			
RT	"Complete genome sequence of the model actinomycete Streptomyces			
RT	coelicolor A3(2)."			
RL	Nature 417:141-147(2002).			
DR	EMBL; AJ007731; CA07627.1;			
DR	EMBL; AL939127; CAB60185.1;			
DR	InterPro; IPR005509; AFSA_repeat.			
DR	Pfam; PF03756; AFSA; 2.			
KM	Complete proteome.			
SO	SEQUENCE 314 AA; 33747 MW; 0F74C9669B6CF389 CRC64;			
Query Match	99.4%; Score 1634.5; DB 2; Length 314;			
Best Local Similarity	99.7%; Pred. No. 2.7e-131;			
Matches 313; Conservative 0; Mismatches 0; Indels 1; Gaps 1;				
OY	1 MPEAVVLINSASDANSIEQTALPVPMALVHRTVRQDAFPVSMIRKGGDRSVTVLPHDH 60			
DB	1 MPEAVVLINSASDANSIEQTALPVPMALVHRTVRQDAFPVSMIRKGGDRSVTVLPHDH 60			
OY	61 PEPAPVHGDRDPLLIETTRQAAALVFHAGYGVPGVGHFLM-TLDTYCHLDHGVSGEV 119			
DB	61 PEPAPVHGDRDPLLIETTRQAAALVFHAGYGVPGVGHFLM-TLDTYCHLDHGVSGEV 120			
OY	120 ALEVEVACSQLKRGQGPVQGVDMVRRAGRLAATGATATRTSPOVYRRMGDFATP 179			

```

Db      121 ALELEAVASQKTFKGGQVQGVDMAYVARRGLATGATATRTSPQVYTRMRGDPATP 180
Qy      180 TASVGTATPVPAARAGRTREDEVDVLSASSQODTWRLRVDTSHPTLFORPNHVPQMLLLE 239
Db      181 TASVGTATPVPAARAGRTREDEVDVLSASSQODTWRLRVDTSHPTLFORPNHVPQMLLLE 240
Qy      240 AARQAACLVTPGAPFVPVPSIGTRFVRYAEFDSPCWIOATVRPAGLTTTAVTGHQGS 299
Db      241 AARQAACLVTPGAPFVPVPSIGTRFVRYAEFDSPCWIOATVRPAGLTTTAVTGHQGS 300
Qy      300 LVFLTLTSGPAPSG 313
Db      301 LVFLTLTSGPAPSG 314

```

## RESULT 2

```

AFSA_STRGR STANDARD; PRT; 301 AA.
ID AFSA_STRGR
AC P18394;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Possible A-factor biosynthesis enzyme.
GN Name=afsa;
OS Streptomyces griseus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxId=1911;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89123125; PubMed=2492509;
RT Horinouchi S., Suzuki H., Nishiyama M., Beppu T.;
RT "Nucleotide sequence and transcriptional analysis of the Streptomyces
RT griseus gene (afsa) responsible for A-factor biosynthesis.";
RT J. Bacteriol. 171:1206-1210(1989).
RN [2]
RN SEQUENCE FROM N.A.
RP Umeyama T.;
RL Submitted (FEB-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: This is a key enzyme for A-factor (2-iscaptyloyl-3R-
CC hydroxymethyl-gamma-butyrolactone) biosynthesis. A-factor is a
CC diffusible bioregulator that is essential for streptomycin
CC production, streptomycin resistance, and spore formation.
CC -----
CC This SWISS-PROT entry is copyrighc. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M24250; AAA65693.1; -
CC DR EMBL; AB011413; BAA32134.1; -.
CC DR PIR; A32061; A32061.
CC DR InterPro; IPR005509; Afsa_repeat.
CC DR Pfam; PF03756; Afsa; 2.
CC KW PfamId.
SQ SEQUENCE 301 AA; 32706 MW; E93302C7547065B3 CRC64;

```

Query Match 62.2%; Score 1023.5; DB 1; Length 301;  
 Best Local Similarity 66.4%; Pred. No. 3.9e-79;  
 Matches 194; Conservative 26; Mismatches 71; Indels 1; Gaps 1;

```

Qy      23 PVPMAVHTRTVQDAFPVSWIKGGRFESVTAVLPDHDFPAPVNGDRDPLLTIAETLRQ 82
Db      10 PVGIEVHTRTREDAPFRPMVRLGRDRFAVEAVLPPDHDFPAPVGGDLPLLVEAMKQ 69
Qy      83 AAMLVHAGYGVGVGHFLMT-LDYTCHDLHGVSGEVALEVEVAQSOLKFRGGQPVQ 141
Db      70 AAMLVHAGYGVGVGHFLMT-LDYTCHDLHGVSGEVALEVEVAQSOLKFRGGQPVQ 129

```

```

Qy      142 QVDMVARRAGRLATGTATRTFTSPQVYTRMRGDFATPTLASVGTAPVPAARAGRTDED 201
Db      130 RVGMAVHRHGDRLATGVAAATFTSTPKAYRRRGVVPVGGSLPETAPVPSAPGARVED 189
Qy      202 VLSASSQODTWRLRVDTSHPTLFORPNHVPQMLLEAARQAACLVTPGAPFVPVPSIGT 261
Db      190 VLSGTRGSEVWEARVDRHPTLFORPNHVPQMLLEAARQAACLVTPGAPFVPVART 249
Qy      262 RFVRYAEFDSPCWIOATVRPAGLTTTAVTGHQGS LVFLTLTSGPAPSG 313
Db      250 RFHRYSEFSPCWIGAVVQGADEDTVTRVTHGHQDETFVSTVLSGPRAHG 301

```

## RESULT 3

```

ID 056152 PRELIMINARY; PRT; 309 AA.
AC 056152;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Jadm1.
GN Name=Jadm1;
OS Streptomyces venezuelae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxId=54571;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ISP5230;
RX MEDLINE=22787062; PubMed=12904539; DOI=10.1099/mic.0.26209-0;
RX Wang L., Vining L.C.;
RT "Control of growth, secondary metabolism and sporulation in
RT Streptomyces venezuelae ISP5230 by Jadm(1), a member of the afsA
RT family of gamma-butyrolactone regulatory genes.";
RT Microbiology 149:1991-2004(2003).
RL EMBL: U24659; AAB36582.2; -.
DR InterPro; IPR005509; Afsa_repeat.
DR Pfam; PF03756; Afsa; 2.
SQ SEQUENCE 309 AA; 33761 MW; 996935C1D855FPAC CRC64;

```

Query Match 36.6%; Score 602; DB 2; Length 309;  
 Best Local Similarity 45.3%; Pred. No. 4e-43;  
 Matches 130; Conservative 39; Mismatches 110; Indels 8; Gaps 3;

```

Qy      29 VHRTRVQDAFPVSWI PKGGRFESVTAVLPDHDFPAPVNGDRDPLLTIAETLRQAMLVF 88
Db      10 VHRADPADIIPTDWTQLKQNFVSARVACLRSLLSRAGARHDPMLVAETIPESTMIVA 69
Qy      89 HAGYGVPGVGHFLM-TLDYTCGDLHGLGSGEVALEVEVAQSOLKFRGGQPVQGVDMAY 147
Db      70 HAEIGVPLDEQFVMMWDLSSADSEALTVDGLSSDVTVVVCSDITTRGSRRLNLTIVVL 129
Qy      148 RPAGRLATGTATRTFTSPQVYTRMRGDFATPTLASVGTAPVPAARAGRTREDEVDV 203
Db      130 TRDRLATGSGTARCTISALAYRRRGKRMALGRPVPLIGVHP--RLVGRARTEDVV 186
Qy      204 LSASSQODTWRLRVDTSHPTLFORPNHVPQMLLEAARQAACLVTPGAPFVPVPSIGTRF 263
Db      187 LAPGRPDQWQLRVNTAHTTLFRPNHVPQMVILLEAARQAATVTTGSAVLPDLSVSF 246
Qy      264 VRYAEFDSPCWIOATVRPAGLTTTAVTGHQGS LVFLTLTSGPA 310
Db      247 LRYELDSPCWIEAESVPTPDSTTTRVTHGHQGS PFVFRCTLSPS 293

```

## RESULT 4

```

ID 024738 PRELIMINARY; PRT; 294 AA.
AC 024738;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE BatX.

```



GN Name=daxh;  
 OS Streptomyces virginiae.  
 OC Bacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycinae; Streptomycetaceae; Streptomyces.  
 NC NCB1\_TaxID=1961;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98037495; PubMed=9371444;  
 RA Kinoshita H., Ipposhi H., Okamoto S., Nakano H., Nishira T., Yamada Y.;  
 RT "Butyrolactone autoregulator receptor protein (BarZ) as a  
 RL transcripional regulator in Streptomyces virginiae."  
 RL J. Bacteriol. 179:6986-6993(1997).  
 DR EMBL; AB001608; BAA23611.1;  
 DR InterPro; IPR005509; AfsA\_repeat.  
 DR Pfam; PF03756; AfsA\_2.  
 SO SEQUENCE 294 AA; 32378 MW; 7192024EB68F7892 CRC64;

Query Match 34.2%; Score 562; DB 2; Length 294;  
 Best local similarity 43.9%; Pred. No. 9.8e-40;  
 Matches 123; Conservative 39; Mismatches 116; Indels 2; Gaps 2;

QY 24 VPMALVHRTYQDAFPVSMIRKGGREFVTAVLPHDHFPAVHGDRHDPILLIAETTRQA 83  
 DB 5 VPRELVHRAAAVEVFLTGMSRTAENRFALTQWPPRAHSYFTPPVNG-CYDPLAESTTKOV 63

QY 84 AMLVHAGYGVVGVYHFLM-TLDYTCGLDHLGVSSEVALEEVVACSQKFRGQGPVQGO 142  
 DB 64 GTLSHAEFGVSRFGQFLMDLHSHVREQAGVGAAAPDLDELVDICSIIRRGRLAKMR 123

QY 143 VDMAVRAAGRLAATGTATRTFTSPQVYRMRGDFATPTASVGTAPVAPARAGRTREDV 202  
 DB 124 YEVLTCGGQYATGCAAFDCTSPAVYQRLRGDRGATGVRLPQPLAPASVGRFLTVDV 183

QY 203 VLSASSQODTRRLRPVDTSHPTLFPQPNHVGEMLLLEAARQAACLVTPAPVBSIGTR 262  
 DB 184 VLSATERLEWQLRDEQHPVLFDPVHVPMVMESARQAQAIDSPRFLPTTMRSE 243

QY 263 FVRVAFDPSPCMIQATVPRGPAAGLTTRVATGHODGSLVF 302  
 DB 244 FSRVAFEDRPFQICQAEPLPAADNGDQVYVGTGHDDITVF 283

RESULT 5  
 Q83X19 PRELIMINARY; PRT; 331 AA.  
 AC Q83X19;  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE A-factor biosynthesis protein AfsA homolog.  
 OS Streptomyces rochei (Streptomyces parvulus).  
 OC plasmid pSLA2-L.  
 OC Bacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycinae; Streptomycetaceae; Streptomyces.  
 NC NCB1\_TaxID=1928;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=7434AN4;  
 RC MEDLINE=22676866; PubMed=12791134;  
 RA Mochizuki S., Hiratsu K., Suwa M., Ishii T., Sugino F., Yamada K.,  
 RA Kinashi H.;  
 RT "The large linear plasmid pSLA2-L of Streptomyces rochei has an  
 RT unusually condensed gene organization for secondary metabolism."  
 RL Mol. Microbiol. 48:1501-1510(2003).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=7434AN4;  
 RC MEDLINE=20408175; PubMed=10954087;  
 RA Hiratsu K., Mochizuki S., Kinashi H.;  
 RT "Cloning and analysis of the replication origin and the telomeres of  
 RT the large linear plasmid pSLA2-L in Streptomyces rochei."  
 RL Mol. Gen. Genet. 263:1015-1021(2000).  
 RN (3)

RP SEQUENCE FROM N.A.  
 RC STRAIN=7434AN4;  
 RX MEDLINE=20231737; PubMed=10767533; DOI=10.1016/S0378-1119(00)00060-3;  
 RA Suwa M., Sugino H., Sasaoka A., Mori E., Fujii S., Shinkawa H.,  
 RA Niimi O., Kinashi H.;  
 RT "Identification of two polyketide synthase gene clusters on the linear  
 RT plasmid pSLA2-L in Streptomyces rochei."  
 RL Gene 246:123-131(2000).  
 RN (4)  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=7434AN4;  
 RC MEDLINE=9803144; PubMed=9836424;  
 RA Kinashi H., Fujii S., Hatani A., Kurokawa T., Shinkawa H.;  
 RT "Physical mapping of the linear plasmid pSLA2-L and localization of  
 RT the eryAI and actI homologs."  
 RL Biosci. Biotechnol. Biochem. 62:1892-1897(1998).  
 DR EMBL; AB088224; BAC76543.1;  
 DR InterPro; IPR005509; AfsA\_repeat.  
 DR InterPro; IPR002086; Aldhyd\_dehydrog.  
 DR InterPro; IPR000623; Shik\_kinase.  
 DR Pfam; PF03756; AfsA\_2.  
 DR PROSITE; PS00687; ALDEHYDE\_DEHYDR\_GLU; UNKNOWN\_1.  
 DR PROSITE; PS01128; SHIKIMATE\_KINASE; UNKNOWN\_1.  
 RW plasmid.  
 SO SEQUENCE 331 AA; 36788 MW; 1E35F5BECFA22068 CRC64;

Query Match 31.9%; Score 524.5; DB 2; Length 331;  
 Best local similarity 38.4%; Pred. No. 1.8e-36;  
 Matches 118; Conservative 44; Mismatches 136; Indels 9; Gaps 5;

QY 10 SASDANSIEQTALEFPMALVHRTVQDAFPVSMIRKGGREFVTAVLPHDHFPAVHGDRHDPILLIAETTRQA 69  
 DB 23 SPQDTSWLTPLTTVPREYVHRAIAEVFLTRCTRIHETRFLLTGCPRAATFTLSPGR 82

QY 70 RHPDLLAETTRQAAMLVHAGYGVVGVYHFLM-TLDYTCGLDHLGVSSEVALEEVVACSQKFRGQGPVQGO 128  
 DB 83 RHPDQIAETTRQVGLHIAHAEFVPLGHHFIMWMSFVSVEHLGVGRPTDLEATVC 142

QY 129 SQLKFRGGQPVQGVQDMAVRAAGRLAATGTATRTFT--SPQVYRMRGDFATPTASVGT 186  
 DB 143 VDVRRRGKLVFELVLTTERDGHVANGG--RFTCTTEAMVRLRSAPATTAAHOAS 200

QY 187 ---APVPAARAGRTREDDVVLSSASSQODTRRLRPVDTSHPTLFPQPNHVGEMLLLEAARQAACLVTPAPVBSIGTR 243  
 DB 201 HQPAPLEPSDGRKARPDDVILAPGAPNRRLNMDTSHPTLFDHSGHVPMTLLESARQ 260

QY 244 AACLVTPG-APVDSIGTRFVRYAEDSPCMIQATVPRGPAAGLTTRVATGHODGSLVF 302  
 DB 261 AACALLPPGSTLIPATVSTERRRYVEFTSPCMIASGLAVSGTFALITGRDDDDVF 320

QY 303 LTTLSGP 309  
 DB 321 TARTSGP 327

RESULT 6  
 Q84H95 PRELIMINARY; PRT; 317 AA.  
 AC Q84H95;  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Transcriptional regulator.  
 OS Streptomyces carzinostaticus subsp. neocarzinostaticus.  
 OC Bacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycinae; Streptomycetaceae; Streptomyces.  
 NC NCB1\_TaxID=167636;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=ATCC 15944;  
 RC Liu W., Nonaka K., Nie L., Bae J., Zazopoulos E., Farnet C.M.,  
 RA Shen B.;  
 RT Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.

DR EMBL, AY117439; AAM78023.1; -.  
 DR InterPro; IPR005509; Alfa\_repeat.  
 DR InterPro; IPR010916; TONB\_Box\_N.  
 DR Pfam; PF03756; Alfa.2.  
 DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; UNKNOWN 1.  
 SQ SEQUENCE 317 AA; 34773 MW; 160B371ABE23B722\_CRC64;

Query Match 31.0%; Score 510; DB 2; Length 317;  
 Best Local Similarity 39.7%; Pred. No. 2.9e-35;  
 Matches 121; Conservative 39; Mismatches 125; Indels 20; Gaps 7;

QY 23 PVPML-----VHRTVQDAFPVSWIPKGGDRPSYAVLPHDPPFAPVHGDRHDL 74  
 DB 10 PVEPALTRTVAKYVHRAALAEVFLTGMMRTGPSFVSAQMPKSHSFSSDRG-VHDL 68  
 QY 75 LIAETLRQAAALVFAAGVGVGVYH-FMTLDYTGCHLHLGVSQVLEVEVACSLK 133  
 DB 69 LICESLRQCGLLTTHAFAVFPVGYQLSWSLSYANPEMFGSPADIEMHVGCTDVR 128  
 QY 134 RGGQPVQGVDMVARRAGLAAVTGATRTFTSPQYRRM--RGDPATPTASVP-GTAPV 189  
 DB 129 RSLFLVAMTMRIEVLVLSLALVAITTFQGHAPAVYRRLRAGRSDAGLCANAPTPAPI 188  
 QY 130 PAARAGRRDEPVVLSASQODTWRLRYDTSHTLFGPNHVPQMLLEAARQAALV- 248  
 DB 189 PYYVGRTRQEDVVLSPEDRLRMRLRVDTNHPVLFDPVDHVPQMLLESVRQGHAVH 248  
 QY 249 TGPAPFVDSIGGTRFVRVAFSDPCMIQATVAPGPAAGLTVRVYGHQDGLVF-----L 303  
 DB 249 PGSGDVPMTLMDVSTNRVYELDEPCMIKEMFSAKARVAGLQGGRLAFAVAVQM 308  
 QY 304 TTLSG 308  
 DB 309 TDTIG 313

## RESULT 7

QY 024740 PRELIMINARY; PRT; 291 AA.  
 AC 024740;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE Fair.  
 GN Name=fair;  
 OS Streptomyces sp.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxId=1931;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FRI-5;  
 RX MEDLINE=97405912; PubMed=9260956;  
 RA Waki M., Mihira T., Yamada Y.,  
 RT "Cloning and characterization of the gene (fair) encoding the receptor  
 RT for an extracellular regulatory factor (Im-2) from Streptomyces sp.  
 RT strain FRI-5.";  
 RL J. Bacteriol. 179:5131-5137 (1997).  
 DR EMBL; AB001683; BAA21858.1; -.  
 DR InterPro; IPR005509; Alfa\_repeat.  
 DR Pfam; PF03756; Alfa.2.  
 SQ SEQUENCE 291 AA; 32146 MW; 9B022EF2F38251C9\_CRC64;

Query Match 29.2%; Score 480.5; DB 2; Length 291;  
 Best Local Similarity 39.9%; Pred. No. 8.8e-33;  
 Matches 116; Conservative 45; Mismatches 121; Indels 9; Gaps 6;

QY 28 LVHRTVQDAFPVSWIPKGGDRPSYAVLPHDPPFAPVHGDRHDLIAETLRQAAALV 87  
 DB 1 VHRRTSTAQVLLTDQRDLDAFVSYARWPLSHAFFTPVGDGVYDPLMCAETIRQLAYLL 60  
 QY 88 FHAGVGVPGVHFMFLMDYTCIL---DHGVSGVALEVEVACSLKRGQGPVQGVGD 144

DB 61 GHAFVAPFGHQFVL-WDLSVSVRPELRLVGLVPAVTDALITCEIKBRAGRLSGLYE 119  
 QY 145 WAVEPAGLAAATGATATRTFTSPQYRRMRGDFE-TPV-ASVPGTAPVPAPAGTRBEDV 202  
 DB 120 AVVRDQGVAVTGAASVTCSTSPAYQRIPEPHVLTPEHRLPLTAPAPAPOSVARSPTDV 179  
 QY 203 VLSASQODTWRLAVDTSHTLFGPNHVPQMLLEAARQAALVGPAPFVSIQGT 262  
 DB 180 VLSPLDRNRVQLVDTNHPVLPFDHWVDHVGWVLMBAARQAASALGRSFMFLGVAGE 239  
 QY 263 FVRVAFSDPCMIQ--TVRPPAAGLTVRVYGHQDGLVFLTTLSGPAF 311  
 DB 240 FKRIVELDAPCVIESERLFDVPGAE-EVVRVYGHQDGLVFLVGTVAASV 289

## RESULT 8

QY 082KU4 PRELIMINARY; PRT; 345 AA.  
 ID 082KU4;  
 AC 082KU4;  
 DT 01-JUN-2003 (TREMBLrel. 24, Created)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE Putative gamma-butyrolactone biosynthesis protein.  
 GN Name=avaA; OrderedLocustNames=SAV2269;  
 OS Streptomyces avermitilis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxId=33903;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MA-4680;  
 RX MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.211433198;  
 RA Omura S., Ikeda H., Ishikawa J., Hamamoto A., Takahashi C.,  
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,  
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;  
 RT "Genome sequence of an industrial microorganism Streptomyces  
 RT avermitilis: deducing the ability of producing secondary  
 RT metabolites.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MA-4680;  
 RX MEDLINE=22608306; PubMed=12692562;  
 RA Ikeda H., Ishikawa J., Hamamoto A., Shinose M., Kikuchi H., Shiba T.,  
 RA Sakaki Y., Hattori M., Omura S.;  
 RT "Complete genome sequence and comparative analysis of the industrial  
 RT microorganism Streptomyces avermitilis.";  
 RL Nat. Biotechnol. 21:526-531 (2003).  
 DR EMBL; AP005030; BAC69980.1; -.  
 DR InterPro; IPR005509; Alfa\_repeat.  
 DR Pfam; PF03756; Alfa.2.  
 KW Complete proteome.  
 SQ SEQUENCE 345 AA; 38292 MW; 7C7FEC7599D70940\_CRC64;

Query Match 19.4%; Score 319.5; DB 2; Length 345;  
 Best Local Similarity 32.2%; Pred. No. 6e-19;  
 Matches 98; Conservative 33; Mismatches 140; Indels 33; Gaps 8;

QY 23 PVPMLVHRTVQDAFPVSWIPKGGDRPSYAVLPHDPPFAPVHGDRHDLIAETLRQ 82  
 DB 28 PVTRELVRSSIAEVLVTDGVRGTGENAFSVGAQMPRDLALHPDENGNDLFLAETLRQ 87  
 QY 83 AAMLVFAAGVGVGVYHFL--MTLDYTGCHLHLGVSQVLEVEV--ACSLKFRGQ 137  
 DB 88 AHFYGAHTYFGVPGVSGRFIGQVSPFLT--DPTALRVGAALVAVLNGTWTEERDRGR 144  
 QY 138 PVQGVDMVARRAGLAAVTGATRTFTSPQYRRMRGDFALPTASVPGTAVPAA----- 192  
 DB 145 PAGARLDVTLTVVDGPPCGRGHTRGIMLDDRYRLRG--RPAAGVSPRPAPADARIA 201  
 QY 193 ---RAGTRDEPVVLSASQODTWRLRYDTSHTLFGPNHVPQMLLEAARQAALV 249  
 DB 202 RPNRVGRRLKMDCVLERPRPDQNRKLRVDRHVAVLFDHPTDHPVLMVLBEGRQGHGLTV 261

Oy	21	ALPVPMLVHRTKRVQDAFPVSMIPKGGDRSSYVALPHDHPFAPVGHDDHDLIAETL	80
Dd	26	ARPLPPADVHKAAAEVLITDAPRLGNRRFAVAALWPRNFFLARATSSPCDPLALETI	85
Oy	81	ROAMLVFNHGUGVGVGYHTFLMTLDYCHLDHGVSGSEVALEVEVACSQLKFGCGPOVQ	140
Dd	86	RQSILHSLSHFCDVPFIHHFVL-----SGLDDLDPVND-----GRLPV	127
Oy	141	GQYDMAVRRAQRGLATGTATT-----RF--TSPOV---	169
Dd	128	-----LDVTSTKTTPRMRARALNDVVYAGLRGRCAIRFEVLAPRRAYMI	175
Oy	170	--RRMGDFATPTTASVPGTAPVPAPARGTRDEDVVL-SASSOODT-WRLRVDTSHPTLF	225
Dd	176	RDRARRARRBPQAQAAAQAATATLPPETVGFHDHLLVLTATQGPDIMQRLRDHPLVF	235
Oy	226	QRPDHYPGMLLEEAQAQACLVTGPAP--FVP-----SIGCRFRVXYAFSPSCWIQAT	278
Dd	236	DHSDHISGMALLBACHQAATATLTPPAAGFGRPQVALTVASSYQAIFGLDSP--VTTT	293
Oy	279	VRP-----GPAAGLTIVRVTHGDGSIVFTTLTSGPAFSG	313
Dd	294	TLPAAHGHSFDSGTRITQLTRAQCSRLLTATVTTTTTAG	333
 RESULT 10 O7NCZ7 PRELIMINARY; PRT; 1199 AA.			
ID	O7NCZ7		
AC	O7NCZ7;		
DT	01-MAR-2004 (TrEMBLrel. 26, Created)		
DT	01-MAR-2004 (TrEMBLrel. 26, Last sequence update)		
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)		
GN	G1128329 protein.		
OS	OrderedLocustNames=g112829;		
OC	Gloeobacter violaceus.		
Bacteria	Cyanobacteria; Chroococcales; Gloeobacter.		
OX	NCBI_Taxid=33072;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
KX	STRAIN=PCC 7421;		
KX	MEDLINE=22977040; PubMed=14621232;		
RA	Nakamura Y., Kaneko T., Sato S., Mimuro M., Miyahita H., Tauchiya T.,		
RA	Sasamoto S., Watanabe A., Kawashima K., Kishida Y., Kiyokawa C.,		
RA	Kohara M., Matsumoto M., Matsuno A., Nakazaki N., Shimpoo S.,		
RT	Takeuchi C., Yamada M., Tabata S.;		
RT	"Complete genome structure of Gloeobacter violaceus PCC 7421, a		
RT	Cyanobacterium that lacks thylakoids";		
RL	DNA Res. 10:137-145 (2003).		
DR	EMBL; AP006578; BAC90770.1; --		
DR	GO; GO:0048037; F:cofactor binding; IEA.		
DR	GO; GO:0016740; F:transferase activity; IEA.		
DR	GO; GO:0006633; P:fatty acid biosynthesis; IEA.		
DR	GO; GO:0008152; P:metabolism; IEA.		
DR	InterPro; IPRO09081; ACP like.		
DR	InterPro; IPRO01227; Ac transferase.		
DR	InterPro; IPRO05509; AfsA_repeat.		
DR	InterPro; IPRO00794; ketoacyl_synth.		
DR	InterPro; IPRO06163; phspanteth bind.		
DR	InterPro; IPRO06162; Pantine_S.		
DR	Pfam; PF00698; Acyl_transf_1; 1.		
DR	Pfam; PF03756; AfsA; 1.		
DR	Pfam; PF00109; ketoacyl_synt; 1.		
DR	Pfam; PF02801; ketoacyl_synt_C; 1.		
DR	Pfam; PF00550; PP-binding; 1.		
DR	PROSITE; PS50075; ACP DOMAIN; 1.		
DR	PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.		
DR	PROSITE; PS00012; PHOSPHOPANTETHINE; UNKNOWN_1.		
KM	Complete proteome, Transference.		
SO	SEQUENCE 1199 AA; 130726 MW; 1554AAE63CA7D543 CRC64;		



```

Db      110 ELDCCSSLDHPTDQKLIPEFIKKIOEASOGKIFGVIP---QYHSSVNSAGSSAPVSTA 166.
Qy      160 TTRFTSPQVYRRMGDPFATPTASVGTAPVPAARAGTRDE-----DVL-----SA 206
Db      167 NSTEDARARA-KNARDHSLSENEKGTGDVCSAPAGRNQSPSSGPRGEPFLAKQPSRP 225
Qy      207 SSQODTWLRVD-----TSHP-----TLFORPNH 231
Db      226 SGEQGGGLSPQGVSKTLDGPESNPLEVHEEPLSGKMEIFTLFNKPKSH 274

```

## RESULT 13

```

Y084_HUMAN STANDARD; PRT; 648 AA.
ID      1084_HUMAN
AC      014659;
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      05-JUN-2004 (Rel. 44, Last annotation update)
DN      Hypothetical protein KIAA0084 (HA2022) (Fragment).
GN      Name=KIAA0084;
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Bone marrow;
RX      MEDLINE=95308325; Pubmed=7788527;
RA      Nagase T., Miyajima N., Tanaka A., Sazuka T., Seki N., Sato S.,
RA      Tabata S., Ishikawa K.-I., Kawarabayashi Y., Kotani H., Nomura N.;
RT      "Prediction of the coding sequences of unidentified human genes. III.
RT      The coding sequences of 40 new genes (KIAA0081-KIAA0120) deduced by
RT      analysis of cDNA clones from human cell line KG-1.";
RL      DNA Res. 2:37-43(1995).

```

```

CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).

```

```

CC      EMBL; D42043; BAA07644.1; -.
KM      Hypothetical protein.
FT      NON TER
SQ      SEQUENCE 648 AA; 70463 MW; 88F68A62D6C1CDA1 CRC64;

```

```

Query Match      7.2%; Score 118.5; DB 1; Length 648;
Best Local Similarity 27.1%; Pred. No. 0.19;
Matches 62; Conservative 22; Mismatches 82; Indels 63; Gaps 11;

```

```

Qy      50 FSVTAVLPHNDPFPAPVAGDRHDP-----LIAETLRQAAMLVFNAQGVGVGYHELM 102
Db      147 FSLAL-----HPQVPTHERKPTLEHIFRAILIKKTDRSQKTDLHNGY-----IL 194
Qy      103 TLDITCHLDHGVSEVAVEL--EVEVACSO-LKRRGGQPVGGQVDMVARRGRILAATTA 159
Db      195 ELDCSSSLDHPDOKLIPEFIKKIOEASOGKIFGVIP---QYHSSVNSAGSSAPVSTA 251
Qy      160 TTRFTSPQVYRRMGDPFATPTASVGTAPVPAARAGTRDE-----DVL-----SA 206
Db      252 NSTEDARARA-KNARDHSLSENEKGTGDVCSAPAGRNQSPSSGPRGEPFLAKQPSRP 310
Qy      207 SSQODTWLRVD-----TSHP-----TLFORPNH 231
Db      311 SGEQGGGLSPQGVSKTLDGPESNPLEVHEEPLSGKMEIFTLFNKPKSH 359

```

```

RESULT 14
O30480 PRELIMINARY; PRT; 1762 AA.
ID      O30480

```

```

AC      O30480;
DT      01-JAN-1998 (TREMELREL. 05, Created)
DT      01-JAN-1998 (TREMELREL. 05, Last sequence update)
DT      01-MAR-2004 (TREMELREL. 26, Last annotation update)
DE      PKs module 2.
OS      Streptomyces hygroscopicus.
OC      Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales;
OC      Streptomycinae; Streptomycetaceae; Streptomyces.
OX      NCBI_TaxID=1912;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=ATCC 29253;
RX      MEDLINE=98085969; Pubmed=9426000; DOI=10.1016/S0378-1119(97)00450-2;
RA      Ruan X., Staasi D., Iax S.A., Katz L.;
RT      "A second type-I PKs gene cluster isolated from Streptomyces
RT      hygroscopicus ATCC 29253, a rapamycin-producing strain.";
RL      Gene 203:1-9(1997)

```

```

DR      EMBL; AF007101; AAC38062.1; -.
DR      PIR; T03222; T03222.
DR      GO; GO:0048037; F:cofactor binding; IEA.
DR      GO; GO:0016491; F:oxidoreductase activity; IEA.
DR      GO; GO:0016740; F:transferase activity; IEA.
DR      GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR      GO; GO:0008152; P:metabolism; IEA.
DR      InterPro; IPR009081; ACP-like.
DR      InterPro; IPR001227; Ac transferase.
DR      InterPro; IPR002198; ADH short.
DR      InterPro; IPR000794; Ketocacyl synth.
DR      InterPro; IPR006163; Phosphateth. bind.
DR      Pfam; PF00698; ACP1 trans. 1; 1.
DR      Pfam; PF00106; adh short; 1.
DR      Pfam; PF00109; ketocacyl-synt; 1.
DR      Pfam; PF02801; ketocacyl-synt C; 1.
DR      Pfam; PF00550; PP-binding; 1.
DR      PROSITE; PS50075; ACP DOMAIN; 1.
DR      PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
KM      Transferrase.
SQ      SEQUENCE 1762 AA; 189185 MW; 15BDE12851743835 CRC64;

```

```

Query Match      6.7%; Score 109.5; DB 2; Length 1762;
Best Local Similarity 24.6%; Pred. No. 3.6;
Matches 93; Conservative 32; Mismatches 162; Indels 91; Gaps 19;

```

```

Qy      10 SASDANSIEOTALEVPMALVHRTFVODAFVSWIPKGDSPSVTAVALPHNDPFPAPV--- 66
Db      697 SIALNPRHATVLSGDRITLHLATQANTKTNML-NVSHA FHSPLMQPIQPFITTLATL 755
Qy      67 -HGRHNDPLLIATETLRQAAMLVFNAQGVGVGYHELMITLDYTCHLDHGVSEVAVELV- 124
Db      756 THNPHPTPLISMATGTPHPTTH-----WTQHITAPVAYTDTLHLHLHNGITVLEIG 809
Qy      125 -EVACSOGLKRRGGQPVGGQVDMVARR-----AGILAATG-----TATTRFT 164
Db      810 PDTLTTLALA-RTTLPTTHHILPTTRRNHNEVRSTNEALGVFSGHSVDWRALPTQRRRT 868
Qy      165 SPQVYRRMGDPFATPTASVGTAPVPAARAGTRDEDEVLSASQODTWLR---RYDTS 220
Db      869 SLPTTPQRDRDFWLHDA-GGAEVGGAGLG-TIDHPLLAATVTVADTGELIISGRISTS 926
Qy      221 -HPTLFORPNH-----VPGMLLEAARQA-----CLVTGPAPFVP 256
Db      927 THNPL-----TDHYNNGVIVYPTGTLIDLAALAEHDTHTVDELVHTPLATLHPTPSQS 982
Qy      257 SIG-----GRFVRYAEFDP-CWIOATV-----RPGPAAGLTTRVYTGQDSGLVLT 305
Db      983 TVGAEITDGRHPLALHSRDVTGTRHTTGLSNQTHPAELSTWSP---DAHQIDLT 1039
Qy      306 L-----SGPAPSG 313
Db      1040 AYQQLADTGLHYGPAPFG 1057

```

```

RESULT 15

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 22:12:24 ; Search time 1184 Seconds  
(without alignment)  
102.621 Million cell updates/sec

Title: US-10-017-471b-17

Perfect score: 1645  
Sequence: 1 MPEAVVLINSASDANSIEQT.....GHQDGLVFTLTSSGPAFSG 313

Scoring table: BLOSUM62  
Gapop 10.0 ; Gapext 0.5

Searched: 1736639 seqs, 388188149 residues

Total number of hits satisfying chosen parameters: 1736639

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubppaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/2/pubppaa/US11\_PUBCOMB.pep.\*
- 20: /cgn2\_6/ptodata/2/pubppaa/US11\_NEW\_PUB.pep.\*
- 21: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep.\*
- 22: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1645	100.0	313	US-10-017-471A-17	Sequence 17, Appl
2	1634.5	99.4	314	US-10-168-663-14	Sequence 14, Appl
3	1023.5	62.2	301	US-10-168-663-13	Sequence 13, Appl
4	562.5	34.2	314	US-10-168-663-15	Sequence 15, Appl
5	482.5	29.3	291	US-10-168-663-16	Sequence 16, Appl
6	319.5	19.4	345	US-10-156-761-9806	Sequence 9806, Ap
7	288	17.5	57	US-10-017-471A-17	Sequence 14, Appl
8	271.5	16.5	353	US-10-168-663-17	Sequence 17, Appl
9	118.5	7.2	648	US-10-755-889-155	Sequence 455, App
10	106.5	6.5	447	US-10-425-115-285291	Sequence 285291,
11	106.5	6.5	460	US-10-425-114-66345	Sequence 66345, A

12	105.5	6.4	256	16	US-10-425-115-204431	Sequence 204431,
13	104.5	6.2	607	14	US-10-156-761-9374	Sequence 9374, Ap
14	101.5	6.2	987	15	US-09-934-070-2	Sequence 2, Appl1
15	101.5	6.2	987	15	US-10-222-772-2	Sequence 2, Appl1
16	101.5	6.2	987	15	US-10-222-772-60	Sequence 60, Appl1
17	101.5	6.2	1002	10	US-09-934-070-4	Sequence 4, Appl1
18	101.5	6.2	1002	15	US-10-222-772-4	Sequence 4, Appl1
19	101.5	6.2	1002	15	US-10-222-772-58	Sequence 58, Appl1
20	101	6.1	3537	18	US-10-922-282-4	Sequence 4, Appl1
21	101	6.1	6396	10	US-09-940-3168-72	Sequence 72, Appl1
22	100.5	6.1	292	14	US-10-156-761-14784	Sequence 11013, A
23	100.5	6.1	777	14	US-10-156-761-14784	Sequence 14784, A
24	100	6.1	751	16	US-10-425-115-218590	Sequence 218590,
25	99	6.0	315	15	US-10-425-114-38543	Sequence 38543, A
26	99	6.0	365	15	US-10-425-114-38900	Sequence 38900, A
27	99	6.0	430	14	US-10-425-114-72573	Sequence 72573, A
28	99	6.0	430	14	US-10-156-761-13294	Sequence 13294, A
29	99	6.0	559	15	US-10-425-114-54109	Sequence 54109, A
30	99	6.0	700	16	US-10-425-115-218593	Sequence 218593,
31	99	6.0	3588	16	US-10-741-601-378	Sequence 378, App
32	99	6.0	3588	17	US-10-741-601-1104	Sequence 1104, App
33	99	6.0	4346	16	US-10-741-601-377	Sequence 377, App
34	99	6.0	4346	17	US-10-741-601-1103	Sequence 1103, App
35	99	6.0	4347	16	US-10-741-601-376	Sequence 376, App
36	99	6.0	4347	17	US-10-741-601-1102	Sequence 1102, App
37	99	6.0	4370	16	US-10-408-765A-1267	Sequence 1267, App
38	97	5.9	4391	16	US-10-478-451-1	Sequence 1, Appl1
39	97	5.9	4393	15	US-10-231-956A-366	Sequence 366, App
40	97	5.9	4393	17	US-10-741-600-1105	Sequence 1105, App
41	96.5	5.9	495	11	US-09-833-245-1680	Sequence 1680, Ap
42	96.5	5.9	495	17	US-10-473-519-16	Sequence 16, Appl1
43	96.5	5.9	495	18	US-10-416-738-19	Sequence 19, Appl1
44	96	5.8	406	15	US-10-344-738-106	Sequence 106, App
45	96	5.8	504	14	US-10-156-761-12417	Sequence 12417, A

ALIGNMENTS

RESULT 1  
US-10-017-471A-17  
Sequence 17, Application US/10017471A  
; Publication No. US20030124644A1  
; GENERAL INFORMATION:  
; APPLICANT: Takano, Eriko  
; TITLE OR INVENTION: Antibiotic Production  
; FILE REFERENCE: 0380-P023290S1  
; CURRENT APPLICATION NUMBER: US/10/017,471A  
; PRIOR FILING DATE: 2001-10-23  
; PRIOR APPLICATION NUMBER: US 60/242,561  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 17  
; LENGTH: 313  
; TYPE: PRT  
; ORGANISM: Streptomyces coelicolor  
US-10-017-471A-17  
Query Match 100.0%; Score 1645; DB 14; Length 313;  
Best Local Similarity 100.0%; Pred. No. 8.4e-154;  
Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MPEAVVLINSASDANSIEQTALPVPMALVHRTYQDAFPVSWIPKGDGRSVTAVALPHDH 60  
DB 1 MPEAVVLINSASDANSIEQTALPVPMALVHRTYQDAFPVSWIPKGDGRSVTAVALPHDH 60  
QY 61 PFPAPVAGDRHDPPLILLETTRQAAALVEHAGYGVGVYHFLMTADYTCCHDLGVSGEVA 120  
DB 61 PFPAPVAGDRHDPPLILLETTRQAAALVEHAGYGVGVYHFLMTADYTCCHDLGVSGEVA 120  
QY 121 ELEVEVACSQLKFRGGOPVOGVDMAVRRAGRLAATGATATRTSPQVYRRMRGDFATPT 180



Db 121 ELEVAVACSQKFRGQPVQGVDMVAVRAGLAATGATATRTFISBPQVRRRGDFAPPT 180  
Qy 181 ASVPETAVPAPARAGRTDEDEVLSASSQODTWRLRVDTSHPTLFQRPNDHVPKMLLEA 240  
Db 181 ASVPETAVPAPARAGRTDEDEVLSASSQODTWRLRVDTSHPTLFQRPNDHVPKMLLEA 240  
Qy 241 AROAACLVTPAPAPVPSIGTRFVRYAEFDSPCWIOATVRPGPAGLTTVAVTGHQDGS 300  
Db 241 AROAACLVTPAPAPVPSIGTRFVRYAEFDSPCWIOATVRPGPAGLTTVAVTGHQDGS 300  
Qy 301 VFLTTLGPAFSG 313  
Db 301 VFLTTLGPAFSG 313

RESULT 2  
US-10-168-663-14  
; Sequence 14, Application US/10168663  
; Publication No. US20040086962A1  
; GENERAL INFORMATION:  
; APPLICANT: Plant Bioscience Limited  
; APPLICANT: Chater, Keith F  
; APPLICANT: Bruton, Celia J  
; APPLICANT: O'Rourke, Sean J  
; APPLICANT: Wietzorek, Andreas W  
; TITLE OF INVENTION: Methods and Materials Relating to Gene Expression  
; FILE REFERENCE: 0380-P02909US0  
; CURRENT APPLICATION NUMBER: US/10/168,663  
; PRIOR FILING DATE: 2002-10-25  
; PRIOR APPLICATION NUMBER: PCT/GB00/04972  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: GB 9930477.6  
; PRIOR FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 14  
; LENGTH: 314  
; TYPE: PRT  
; ORGANISM: Streptomyces sp.  
US-10-168-663-14

Query Match 99.4%; Score 1634.5; DB 15; Length 314;  
Best Local Similarity 99.7%; Pred. No. 9,2e-153;  
Matches 313; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
Qy 1 MBEAVVLINSASDANSIEGTALPVPMAIVHRRVODAFPVSWIPKGGDRFSVTAVLPHDH 60  
Db 1 MBEAVVLINSASDANSIEGTALPVPMAIVHRRVODAFPVSWIPKGGDRFSVTAVLPHDH 60  
Qy 61 PFFAPVHGRHDPFLIAETLRQAAMLVFAGYGVPGYHFLM-TLDYTCMLDHLGVSCEV 119  
Db 61 PFFAPVHGRHDPFLIAETLRQAAMLVFAGYGVPGYHFLM-TLDYTCMLDHLGVSCEV 120  
Qy 120 ALEVEVACSQKFRGQPVQGVDMVAVRAGLAATGATATRTFISBPQVRRRGDFAPPT 179  
Db 121 ALEVEVACSQKFRGQPVQGVDMVAVRAGLAATGATATRTFISBPQVRRRGDFAPPT 180  
Qy 180 TASVGTAPVAPARAGRTDEDEVLSASSQODTWRLRVDTSHPTLFQRPNDHVPKMLLE 239  
Db 181 TASVGTAPVAPARAGRTDEDEVLSASSQODTWRLRVDTSHPTLFQRPNDHVPKMLLE 240  
Qy 240 AROAACLVTPAPAPVPSIGTRFVRYAEFDSPCWIOATVRPGPAGLTTVAVTGHQDGS 299  
Db 241 AROAACLVTPAPAPVPSIGTRFVRYAEFDSPCWIOATVRPGPAGLTTVAVTGHQDGS 300  
Qy 300 LVFLTTLGPAFSG 313  
Db 301 LVFLTTLGPAFSG 314

RESULT 3  
US-10-168-663-13

; Sequence 13, Application US/10168663  
; Publication No. US20040086962A1  
; GENERAL INFORMATION:  
; APPLICANT: Plant Bioscience Limited  
; APPLICANT: Chater, Keith F  
; APPLICANT: Bruton, Celia J  
; APPLICANT: O'Rourke, Sean J  
; APPLICANT: Wietzorek, Andreas W  
; TITLE OF INVENTION: Methods and Materials Relating to Gene Expression  
; FILE REFERENCE: 0380-P02909US0  
; CURRENT APPLICATION NUMBER: US/10/168,663  
; PRIOR FILING DATE: 2002-10-25  
; PRIOR APPLICATION NUMBER: PCT/GB00/04972  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: GB 9930477.6  
; PRIOR FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 13  
; LENGTH: 301  
; TYPE: PRT  
; ORGANISM: Streptomyces sp.  
US-10-168-663-13

Query Match 62.2%; Score 1023.5; DB 15; Length 301;  
Best Local Similarity 66.4%; Pred. No. 2,2e-92;  
Matches 194; Conservative 26; Mismatches 71; Indels 1; Gaps 1;  
Qy 23 PVPMAIVHRRVODAFPVSWIPKGGDRFSVTAVLPHDHPPAPVHGRHDPFLIAETLRQ 82  
Db 10 PVEIEMVHRTPEDEAFPPANWRGLGRDRFAVAALPHDHFFAPVGGDLHDHLVAEAMRQ 69  
Qy 83 AAMLVFAGYGVPGYHFLM-TLDYTCMLDHLGVSCEV ALEVEVACSQKFRGQPVQGV 141  
Db 70 AAMLVFAGYGVPGYHFLM-TLDYTCMLDHLGVSCEV ALEVEVACSQKFRGQPVQGV 129  
Qy 142 QVDMVAVRAGLAATGATATRTFISBPQVRRRGDFAPPTASVPETAVPAPARAGRTDED 201  
Db 130 RVGMAVHGRHDPFLIAETLRQAAMLVFAGYGVPGYHFLM-TLDYTCMLDHLGVSCEV 189  
Qy 202 VVLSSSQODTWRLRVDTSHPTLFQRPNDHVPKMLLEAARQAACLVTPAPVPSIGT 261  
Db 190 VVLSSSQODTWRLRVDTSHPTLFQRPNDHVPKMLLEAARQAACLVTPAPVPSIGT 249  
Qy 262 RFVRYAEFDSPCWIOATVRPGPAGLTTVAVTGHQDGS LVFLTTLGPAFSG 313  
Db 260 RFVRYAEFDSPCWIOATVRPGPAGLTTVAVTGHQDGS LVFLTTLGPAFSG 301

RESULT 4  
US-10-168-663-15  
; Sequence 15, Application US/10168663  
; Publication No. US20040086962A1  
; GENERAL INFORMATION:  
; APPLICANT: Plant Bioscience Limited  
; APPLICANT: Chater, Keith F  
; APPLICANT: Bruton, Celia J  
; APPLICANT: O'Rourke, Sean J  
; APPLICANT: Wietzorek, Andreas W  
; TITLE OF INVENTION: Methods and Materials Relating to Gene Expression  
; FILE REFERENCE: 0380-P02909US0  
; CURRENT APPLICATION NUMBER: US/10/168,663  
; PRIOR FILING DATE: 2002-10-25  
; PRIOR APPLICATION NUMBER: PCT/GB00/04972  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: GB 9930477.6  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 15  
; LENGTH: 314  
; TYPE: PRT  
; ORGANISM: Streptomyces sp.

US-10-168-663-15

Query Match 34.2%; Score 562.5; DB 15; Length 314;  
 Best Local Similarity 42.9%; Pred. No. 8.7e-47;  
 Matches 126; Conservative 39; Mismatches 118; Indels 11; Gaps 3;

QY 19 QTALE-----VPMALHRTVQDAFPVSWIPKGGDSVTAVLPHDHPFAPVHGD 69  
 DB 11 QTGPRPASAMTSTVPRELVHRAVAEVLFGMSRTAENRFPALTAQMPRAHSYFTPVNG- 69  
 QY 70 RHDPLLAETLRQAMLVFAHAGVGPVGHFLM-TLDYTCGLDHLGSGVAELLEVAC 128  
 DB 70 CYDPLASEITRQVGLTLSHAFVSFGDQPLMDLHLSVRPEQAGVGAAPADLELVIC 129  
 QY 129 SOLKFRGGQPVQGVDAVARRAGLAATGTATRTFTSPQVYRRMGDFATPTASVPGTAP 188  
 DB 130 SDIRRRGRRLAGMYETLYGGQVATGGAAPCTSPAYVQRGRGVATGVRLPQP 189  
 QY 189 VPAARAGRTDEDEVLASQODTWRLRVDTSHPTLFQRPNDHVPKMLLEAARQAACLV 248  
 DB 190 LAPASVGRFLTTVDVLATERPLEMQLRVDEQHPVLFQHPVDHVPKMLMESARQAQAI 249  
 QY 249 TGPAPFPISIGTFRVYAEFDSQVQATVPBPAGLTVRTYTGQDGLVF 302  
 DB 250 DPNRPLPTWRSEFSRYAEIDRCPWIOAEPLPADNGDRQVRYGHQDDTTVF 303

RESULT 5

US-10-168-663-16  
 ; Sequence 16, Application US/10168663  
 ; Publication No. US20040086962A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Plant Bioscience Limited  
 ; APPLICANT: Chater, Keith F  
 ; APPLICANT: Bruton, Ceilia J  
 ; APPLICANT: O'Rourke, Sean J  
 ; APPLICANT: Metcortek, Andreas W  
 ; TITLE OF INVENTION: Methods and Materials Relating to Gene Expression  
 ; FILE REFERENCE: 0380-80290US0  
 ; CURRENT APPLICATION NUMBER: US/10/168, 663  
 ; PRIOR FILING DATE: 2002-10-25  
 ; PRIOR APPLICATION NUMBER: PCT/GB00/04972  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: GB 9930477.6  
 ; NUMBER OF SEQ ID NOS: 35  
 ; SOFTWARE: Patent version 3.0  
 ; SEQ ID NO 16  
 ; LENGTH: 291  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces sp.  
 US-10-168-663-16

Query Match 29.3%; Score 482.5; DB 15; Length 291;  
 Best Local Similarity 40.2%; Pred. No. 6.3e-39;  
 Matches 117; Conservative 44; Mismatches 121; Indels 9; Gaps 6;

QY 28 LVHRTVQDAFPVSWIPKGGDSVTAVLPHDHPFAPVHGDRLDPLLAETLRQAMLV 87  
 DB 1 LVHRTSTPAQVLLTQMORLDDARFSTYARWPLSHAFETPVGDGYDPLMCAETIRQIAYLL 60  
 QY 88 FHAGVGVPHFLMTDYLCHL---DHLGSGVAELLEVACQKFRGGQPVQGVQVD 144  
 DB 61 GHAEPAVFGHQFVL-WDLSSVVRPELRLGLVPAVDLITCEVIRRRGRSLGAGYE 119  
 QY 145 WAVEARAGLAATGATATRTFTSPQVYRRMGDFATPTASVPGTAPVPAARAGRTDEVD 202  
 DB 120 AVARRDQGVATGASVTCISPAVYQRIRPREHVLTPERRPLPLAPAPQVVALSPDV 179  
 QY 203 VLSASSQODTWRLRVDTSHPTLFQRPNDHVPKMLLEAARQAACLVTPAPFPVPSIGTR 262  
 DB 180 VLSPLDENRQQLRVDTSHPTLFQRPNDHVPKMLLEAARQAACLVTPAPFPVPSIGTR 239

QY 263 FVRYAEFDSQVQAT--TVRGPAGLTVRYTGQDGLSVLELTLSGPAF 311  
 DB 240 FKRYVELDACVLESRLFQDVPGE--EVRYTGQDGLSVLELTLSGPAF 289

RESULT 6

US-10-156-761-9806  
 ; Sequence 9806, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156, 761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 9806  
 ; LENGTH: 345  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces avermitilis  
 US-10-156-761-9806

Query Match 19.4%; Score 319.5; DB 14; Length 345;  
 Best Local Similarity 32.2%; Pred. No. 1e-22;  
 Matches 98; Conservative 33; Mismatches 140; Indels 33; Gaps 8;

QY 23 PVPALVHRTVQDAFPVSWIPKGGDSVTAVLPHDHPFAPVHGDRLDPLLAETLRQ 82  
 DB 28 PVRRELVRSSIAEFTVDTGRTGENAFVGAQMPRDHALYHPDENGLNDPLFAETLRQ 87  
 QY 83 AMLVFAAGYGVPHFL---MTLDYTCGLDHLGSGVAELLEVAC--ACSQLKFRGGQ 137  
 DB 88 AHFGATYFVGVPVSGFFIQDVSFEIT--DPTALVGAAPLAVLVNGTWTERDRGR 144  
 QY 138 PVQGVQVMAVRBRAGLAATGTATRTFTSPQVYRRMGDFATPTASVPGTAPVPA----- 192  
 DB 145 PAGARLDVTLTVDRPCGGRHTGMLMDRRYRLRG--RPAASGVSPRRAPADARIA 201  
 QY 193 ---RAGRTDEDEVLASQODTWRLRVDTSHPTLFQRPNDHVPKMLLEAARQAACLV 249  
 DB 202 RPNRVRGLRKDCVLEDRDQDWRRLRVDDHVLPHPTDHPVLMVMBGFRGLHLY 261  
 QY 250 GPAPFPVPSIGTRPVRYAEFDSPC-----WQATVRGPAGLTT---VRVGHOD 297  
 DB 262 HEAS--RRTIGDBAFALAGSLDCAAFGEIGETILSLKGPSEGVPTBCALRVAAHQG 319  
 QY 298 GSLV 301  
 DB 320 ERL 323

RESULT 7

US-10-017-471A-14  
 ; Sequence 14, Application US/10017471A  
 ; Publication No. US2003012464A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Takano, Eriko  
 ; APPLICANT: Bibb, Merwyn  
 ; TITLE OF INVENTION: Antibiotic Production  
 ; FILE REFERENCE: 0380-P023290S1  
 ; CURRENT APPLICATION NUMBER: US/10/017, 471A  
 ; CURRENT FILING DATE: 2001-10-23  
 ; PRIOR APPLICATION NUMBER: US 60/242,561

```

; PRIOR FILING DATE: 2000-10-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 14
; LENGTH: 57
; TYPE: PRT
; ORGANISM: Streptomyces coelicolor
US-10-017-471A-14

```

```

Query Match      17.5%; Score 288; DB 14; Length 57;
Best Local Similarity 100.0%; Pred. No. 1.1e-20;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db      1 MPEAVVILNSASDANSIECTALPVPMALVHRVRVDAPFVSNIPKGGDFSTYATLP 57
      1 MPEAVVILNSASDANSIDQTLAPVPMALVHRVRVDAPFVSNIPKGGDFSTYATLP 57

RESULT 8
US-10-168-663-17
; Sequence 17, Application US/10168663
; Publication No. US20040086962A1
; GENERAL INFORMATION:
; APPLICANT: Plant Bioscience Limited
; APPLICANT: Chater, Keith F
; APPLICANT: Bruton, Celia J
; APPLICANT: O'Rourke, Sean J
; APPLICANT: Wietzorek, Andreas W
; TITLE OF INVENTION: Methods and Materials Relating to Gene Expression
; FILE REFERENCE: 0380-P02909US0
; CURRENT APPLICATION NUMBER: US/10/168,663
; CURRENT FILING DATE: 2002-10-25
; PRIOR APPLICATION NUMBER: PCT/GB00/04972
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: GB 9930477.6
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 17
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Streptomyces coelicolor
US-10-168-663-17

```

```

Query Match      16.5%; Score 271.5; DB 15; Length 353;
Best Local Similarity 27.9%; Pred. No. 6e-18;
Matches 95; Conservative 43; Mismatches 123; Indels 79; Gaps 12;

QY      21 ALPVMALVHRVRVDAPFVSNIPKGGDFSTYATLPVPHDHPFAVHGDRHPLIAETL 80
      26 APPLPADVHKAAMAEVLLTDARPGENRFAVAALMPRTFLAHRATSSPCDPLAAETI 85
      81 ROAAMLVFAGVGVFVFLMTLDYTCYLDHLSGSEVAELEVACGSLKFRGQOPVQ 140
      86 ROSAHLSTFCDVPIGHFV-----SGLDLDLDLVWMS-----GPLFV 127
      141 GQDVAVRRAGRLAATGTATT-----RP--TSPQVY-- 169
      128 -----LDVTSTTTNNPRMARALNADYVAGLHGRCAIREVLAPRRYMI 175
      170 --RRNRGDFATPTASVGTAPVAPABAGTRDEYVL--SASSQOQT--WRLRYDTSHPTLF 225
      176 RDRARRAERPAQOAAAGATATLPETVGFHDHVLVLAQGLPTAMQRLRRDHPVLF 235
      226 QRPNDHVRPMLLEAAROAACLVGTGAP--FVP-----SIGTRFVRYAEPDPCMIQAT 278
      236 DHESDHISGMLLEACROATATLTPPAGAFGRVOVALVASSYQAFGEIDSP--VITI 293
      279 VRP-----GPAAGLTVRVYTGHDGSLVFLTTLGPAFSG 313
      294 TLPAHGHSPDSGTRTLQLTARQSGRTLTATVTTTTTAG 333

```

```

RESULT 9
US-10-755-889-455
; Sequence 455, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-KB
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: Patent In version 3.2
; SEQ ID NO 455
; LENGTH: 648
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-755-889-455

```

```

Query Match      7.2%; Score 118.5; DB 16; Length 648;
Best Local Similarity 27.1%; Pred. No. 0.018;
Matches 62; Conservative 22; Mismatches 82; Indels 63; Gaps 11;

QY      50 FSVTAIVLPHDHPFAPVHGDRHPL-----LIAETLROAAMLVFAGVGVVGHFLM 102
      147 FSLAAL-----HPVQPTHEREKTELEHIFRAILIKTIDRSQKTLHNHGYS-----IL 194
      103 TLDTYCHLDHLSGSEVAEL--EVENACSC--LKRGGQOPVGVDMVAPRAGRLAATGTA 159
      195 ELDCSSLDHPTDQKLTFEPIKIQEASQGLKFFVGIYF--QYHSSVNSAGSSAPVSTA 251
      160 TTRFTSPQVYRRMGDFATPTASVGTAPVAPABAGTRDE-----DVVL-----SA 206
      252 NSTEDARDA-KNARGDASLENEKPGTCDVCSAAGNQSPESSGRGVPLAKQSSP 310
      207 SSQDDTWRLRYD-----TSHP-----TLFQRPNDH 231
      311 SGBDGGGLSPQGVSKTLDGPESNPLFVHEEPLSGKMEIFLTFKPKFSH 359

```

```

RESULT 10
US-10-425-115-285291
; Sequence 285291, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovacic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 285291
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: M74577_23284C.1.pep
US-10-425-115-285291

Query Match      6.5%; Score 106.5; DB 16; Length 447;
Best Local Similarity 22.6%; Pred. No. 0.17;
Matches 74; Conservative 23; Mismatches 105; Indels 125; Gaps 14;

QY      62 FPAVHGDRHPLIAETL--ROAAMLVFAGVGVVGHFLMTLDYTC-----LDH 112

```

Db 14 PFLPFGCHISPMQLAHLARGLAVTLHTGNAP-----DATRIHELTVPDIE 64  
 Qy 113 LGVSGEVALEVEVACSQK-----FRGGQP-----VQGVDAVARRAG 151  
 Db 65 SSFPEVTSIGTDIVTQLALNACAPFREALLSLRGGQDVACAVVDDGCVSALRAH 124  
 Qy 152 RLAA-----TGATTTFTSPQVYRRMGDFATPTASVPGTAPPAARAGTRD----- 199  
 Db 125 RLGVPAVLRTDSAAAT-FSSMLAIPRLRDAGFVPEKEERLDEPVDLERLARDLIRVDG 183  
 Qy 200 -----EDVLSASS-----QODTWRLAVDTSHPT-----LFOR 227  
 Db 184 SDTDALCGFIARVADAVARASGVVINTFERMEASELAKIQRELSPAPFVGLHLISQA 243  
 Qy 228 P-----NDHVPGLLLEAARQAACLVTPAPFVPSIGTFVRYAEFDS 271  
 Db 244 PAEGSLAPDRGCLAMLDHPRSVLVYSLSGSAVCVRG-----AFVEMA----- 288  
 Qy 272 PCWIOA-----TVRPGPAAGLTVV 290  
 Db 289 --WQARSQVSLVWVRPGLVGVPEV 313

## RESULT 11

US-10-425-114-66345  
 ; Sequence 66345, Application US/10425114  
 ; Publication No. US20040034888A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Liu, Jingdeng  
 ; APPLICANT: Zhou, Yinhua  
 ; APPLICANT: Kovacic, David K.  
 ; APPLICANT: Screen, Steven E  
 ; APPLICANT: Tabaska, Jack E  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 ; FILE REFERENCE: 38-21(5313)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,114  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 73128  
 ; SEQ ID NO 66345  
 ; LENGTH: 460  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: LIB4729-040-F10\_F11.pep  
 US-10-425-114-66345

Query Match 6.5%; Score 106.5; DB 15; Length 460;  
 Best Local Similarity 22.6%; Pred. No. 0.18;  
 Matches 74; Conservative 23; Mismatches 105; Indels 125; Gaps 14;  
 Qy 62 PFAVHGDRHDPDLIAETL--RQAMLVFNAGVVPVGHFLMTLDYTCR-----LDH 112  
 Db 27 PFLPFGCHISPMQLAHLARGLAVTLHTGNAP-----DATRIHELTVPDIE 77  
 Qy 113 LGVSGEVALEVEVACSQK-----FRGGQP-----VQGVDAVARRAG 151  
 Db 78 SSFPEVTSIGTDIVTQLALNACAPFREALLSLRGGQDVACAVVDDGCVSALRAH 137  
 Qy 152 RLAA-----TGATTTFTSPQVYRRMGDFATPTASVPGTAPPAARAGTRD----- 199  
 Db 138 RLGVPAVLRTDSAAAT-FSSMLAIPRLRDAGFVPEKEERLDEPVDLERLARDLIRVDG 196  
 Qy 200 -----EDVLSASS-----QODTWRLAVDTSHPT-----LFOR 227  
 Db 197 SDTDALCGFIARVADAVARASGVVINTFERMEASELAKIQRELSPAPFVGLHLISQA 256  
 Qy 228 P-----NDHVPGLLLEAARQAACLVTPAPFVPSIGTFVRYAEFDS 271  
 Db 257 PAEGSLAPDRGCLAMLDHPRSVLVYSLSGSAVCVRG-----AFVEMA----- 301  
 Qy 272 PCWIOA-----TVRPGPAAGLTVV 290

Db 302 --WQARSQVSLVWVRPGLVGVPEV 326

## RESULT 12

US-10-425-115-204431  
 ; Sequence 204431, Application US/10425115  
 ; Publication No. US20040214272A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovacic, David K.  
 ; APPLICANT: Zhou, Yinhua  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants  
 ; FILE REFERENCE: 38-21(5322)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,115  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 369326  
 ; SEQ ID NO 204431  
 ; LENGTH: 256  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: MFT4577\_118032C.1.pep  
 US-10-425-115-204431

Query Match 6.4%; Score 105.5; DB 16; Length 256;  
 Best Local Similarity 26.0%; Pred. No. 0.099;  
 Matches 57; Conservative 19; Mismatches 58; Indels 85; Gaps 13;

Qy 112 HLGVSGEVALEVEVACS-----QLKFRGGQPVOGO--VDMVARRA--GRLATGATTTFT 164  
 Db 95 HSNVNRGGRVRNPIVCAAVFQLTRELGLKSDGQITLMLRQAAPSILATSSGTT--- 151  
 Qy 165 SPOVYRRMGDFATPTASVPGTAPVPAARAGTRDEDVLSASSQODTWRLAVDTSHPT 224  
 Db 152 -PAVF-----SCSSAPSTS---AASAGAGAGAVSL-----PLT 181  
 Qy 225 FORP-NDHVPGLLLEAARQAACLVTPAPFVPSIGTFVRYAEFDSPCMIOATVRRP 283  
 Db 182 GKRPREEHEP-----AVAVAVAPF-----MATIQARFVA 212  
 Qy 284 AA-GLT-----TVRVTHGDSLVFLTTLSGPA 310  
 Db 213 AAMGLSPAQEAAQAAYASVAAQGH---LNLISVLSGA 248

## RESULT 13

US-10-156-761-9374  
 ; Sequence 9374, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 9374  
 ; LENGTH: 607  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces avermitilis

US-10-156-761-9374

Query Match 6.4%; Score 104.5; DB 14; Length 607;  
Best Local Similarity 24.0%; Pred. No. 0.41;  
Matches 88; Conservative 28; Mismatches 136; Indels 115; Gaps 17;

QY 3 EAVVLINSASDANSIE-OTALPVMA-----LVHRTYVD----- 36  
DB 175 EAAVLAAAPDSAPBEPAAALAVPOAGDGLVLRVVPASLDDYRAHGGYAAALRA 234  
QY 37 -----APVSMIPKGDPSVTAVLPHPDHPFAPVHGDHR-- 71  
DB 235 FELGPAEIVREYTDAGLVGRGAAPFT-----GKMQATTSRP--DRPHALVCANDESEP 287  
QY 72 -----DPLIAETLRQAAMLVFHAGYGVVGHFLMTLDYTGHLGLVSGEVA 120  
DB 288 GTFKDRVIMEGDPVALVESMTIAAVAV-----GAHQGLYLRG--EYPRALSRM----- 334  
QY 121 ELEVEVACSOLKFRG--GQPVQGG--VDMAVRARGRLAATGTATRTFSQVYRR----- 171  
DB 335 -----SHAEOARARGLDDVYLGQGYARDIEIRGAGAYICGEETALFNISIEYRGEPRS 390  
QY 172 -----MRGDFATPTA--SVPGTAPV--PAARAGRTREDVYLSASSQODTWRLRYDTSHP 222  
DB 391 KRPPEVEKGLRGRPTAENNVETLVNVLVWGAQAYAAIGTAKSTGPKLFCVSGSVDRP 450  
QY 223 TLFQRPNDRVPGMLLEAA--RQAACLVTPGAPFPVPSIGSTRFVVAEPDPSPCWIOATV 279  
DB 451 GVVLELPGATLGGELLELAGVERLRAVLLGGAA-----GG--FVRPDERDIPLTFEGSTR 502  
QY 280 RPPGPAAG 286  
DB 503 QAGTTLG 509

RESULT 14

US-09-934-070-2  
; Sequence 2, Application US/09934070  
; Publication No. US20030092004A1  
; GENERAL INFORMATION:  
; APPLICANT: Lipton, Stuart A.  
; APPLICANT: Zhang, Dongxian  
; APPLICANT: Chatterton, Jon E.  
; APPLICANT: Awobuluyi, Marc  
; APPLICANT: Sevarino, Kevin A.  
; TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS  
; FILE REFERENCE: P-LJ 4900  
; CURRENT APPLICATION NUMBER: US/09/934,070  
; CURRENT FILING DATE: 2001-08-20  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 987  
; TYPE: PRT  
; ORGANISM: Rattus sp.  
US-09-934-070-2

Query Match 6.2%; Score 101.5; DB 10; Length 987;  
Best Local Similarity 23.8%; Pred. No. 1.6;  
Matches 70; Conservative 31; Mismatches 96; Indels 97; Gaps 13;

QY 1 MPEAVVLINSASDANSIEQTALPVPMALVHRTVQDAFPVSMIPKG-----GDRFSVTA 54  
DB 217 VPAAVLLGCGSTARAHEVLEAARPGOWILGTPLPAEALPTTGLPFGVALAGETEOHSLA 276  
QY 55 VLPHDHPFAVHGRHDP--LIAETLRQAAMLVFHAGYGVVGHFLMTLDYTGHL 113  
DB 277 VV-----HDMVELVAQALSSMAL----- 294  
QY 114 GVSGEVALEVEVACSOLK-----FRGGQPVQGVDMAVRARGRLAATGTAT 160  
DB 295 -VHERALIPAVVNCDDLTGSGSEATGRTIARFLGNTSFGQ-----RTGAVVWYTGSSQ 346

QY 161 TRFTSP-QYVRMRMGDFATPTASVPGTAP-----VPAARAGRTREDVYLSASSQODTW 213  
DB 347 VHSVRHFKVMSLRDPLGAPAMATVGSWQOGQDLFOGGAALR-----VPSPGTARP 400  
QY 214 RLRYDT--SHPTLFRPNDR-----VPGMLLY-----BAAR--QAACLVTPGAP 253  
DB 401 KLRVVTIVHHPFVTRSDSDGQCPAGQLCLDPOTNDSALDLALFALVNGSV 454

RESULT 15

US-10-222-772-2  
; Sequence 2, Application US/10222772  
; Publication No. US2004003500A1  
; GENERAL INFORMATION:  
; APPLICANT: Lipton, Stuart A.  
; APPLICANT: Zhang, Dongxian  
; APPLICANT: Chatterton, Jon E.  
; APPLICANT: Awobuluyi, Marc  
; APPLICANT: Sevarino, Kevin A.  
; TITLE OF INVENTION: EXCITATORY GLYCINE RECEPTORS AND METHODS  
; FILE REFERENCE: P-LJ 5338  
; CURRENT APPLICATION NUMBER: US/10/222,772  
; CURRENT FILING DATE: 2002-08-16  
; PRIOR APPLICATION NUMBER: US 09/934,070  
; PRIOR FILING DATE: 2001-08-20  
; NUMBER OF SEQ ID NOS: 62  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 987  
; TYPE: PRT  
; ORGANISM: Rattus sp.  
US-10-222-772-2

Query Match 6.2%; Score 101.5; DB 15; Length 987;  
Best Local Similarity 23.8%; Pred. No. 1.6;  
Matches 70; Conservative 31; Mismatches 96; Indels 97; Gaps 13;

QY 1 MPEAVVLINSASDANSIEQTALPVPMALVHRTVQDAFPVSMIPKG-----GDRFSVTA 54  
DB 217 VPAAVLLGCGSTARAHEVLEAARPGOWILGTPLPAEALPTTGLPFGVALAGETEOHSLA 276  
QY 55 VLPHDHPFAVHGRHDP--LIAETLRQAAMLVFHAGYGVVGHFLMTLDYTGHL 113  
DB 277 VV-----HDMVELVAQALSSMAL----- 294  
QY 114 GVSGEVALEVEVACSOLK-----FRGGQPVQGVDMAVRARGRLAATGTAT 160  
DB 295 -VHERALIPAVVNCDDLTGSGSEATGRTIARFLGNTSFGQ-----RTGAVVWYTGSSQ 346  
QY 161 TRFTSP-QYVRMRMGDFATPTASVPGTAP-----VPAARAGRTREDVYLSASSQODTW 213  
DB 347 VHSVRHFKVMSLRDPLGAPAMATVGSWQOGQDLFOGGAALR-----VPSPGTARP 400  
QY 214 RLRYDT--SHPTLFRPNDR-----VPGMLLY-----BAAR--QAACLVTPGAP 253  
DB 401 KLRVVTIVHHPFVTRSDSDGQCPAGQLCLDPOTNDSALDLALFALVNGSV 454

Search completed: July 20, 2005, 22:42:56  
Job time : 1166 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: July 20, 2005, 16:48:47 ; Search time 120 Seconds  
(without alignments)  
1008,800 Million cell updates/sec

Title: US-10-017-471B-17

Perfect score: 1645  
Sequence: 1 MPRAVYLINSASDANSIEQT.....GHODGSLVLTLSGPAFSG 313

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

1: Geneseq\_16Dec04:\*  
2: geneseqp1980s:\*  
3: geneseqp1990s:\*  
4: geneseqp2000s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1645	100.0	313	5	ABB80941 S. coelic
2	1645	100.0	313	7	ADE14792 Streptomy
3	524.5	31.9	331	8	ADH19754 Streptomy
4	288	17.5	57	5	ABB80938 S. coelic
5	271.5	16.5	353	4	AAU04042 Streptoco
6	126	7.7	541	7	ADCC1753 Human nov
7	118.5	7.2	516	8	ADCC1754 Human nov
8	118.5	7.2	578	8	ABM80427 Tumour-as
9	118.5	7.2	648	8	ADRI4454 Human NF-
10	110	6.7	749	7	ABO72006 Pseudomon
11	104	6.3	660	7	ABO84125 Pseudomon
12	103	6.3	816	7	ABO74546 Pseudomon
13	101.5	6.2	987	7	ADCC7286 Rat NR3B
14	101.5	6.2	987	7	ADCC7286 Rat NR3B
15	101.5	6.2	987	8	ADN49237 Rat N-met
16	101.5	6.2	987	8	ADN49180 Rat N-met
17	101.5	6.2	1002	7	ADCC7284 Rat NR3B
18	101.5	6.2	1002	7	ADCC7231 Rat NR3B
19	101.5	6.2	1002	8	ADN49182 Rat N-met
20	101.5	6.2	1002	8	ADN49235 Rat N-met
21	101	6.1	250	4	AAU63992 Propionib
22	101	6.1	250	4	ABM60511 Propionib
23	100	6.1	1257	6	ABO68544 Pseudomon
24	99.5	6.0	632	7	ABO68402 Pseudomon
25	99	6.0	1142	4	ABG24027 Novel hum

26	99	6.0	3588	8	ADQ39441 Human myo
27	99	6.0	4346	8	ADQ39440 Human myo
28	99	6.0	4347	8	ADQ39439 Human myo
29	99	6.0	4370	7	ADJ69461 Human hea
30	98.5	6.0	507	5	ADRA1466 Human CD-
31	98.5	6.0	1131	4	ABG21347 Novel hum
32	98	6.0	431	4	ABG20554 Novel hum
33	97.5	5.9	388	7	ABO69171 Pseudomon
34	97	5.9	4391	6	AAE34390 Human per
35	97	5.9	4393	4	AAE31889 Human aci
36	97	5.9	4393	8	ADL35758 Human per
37	97	5.9	4393	8	ADQ39442 Human myo
38	97	5.9	4336	4	ABG23265 Novel hum
39	96.5	5.9	495	5	ABG66664 Human alp
40	96.5	5.9	495	5	ABG64931 Human alb
41	96.5	5.9	495	5	ABG77002 Human pro
42	96.5	5.9	495	6	ABM84664 Human SEC
43	96.5	5.9	495	8	ADL78198 Albumin f
44	96	5.8	4881	3	ABE23751 S. averm
45	96	5.8	4881	4	AAE5266 Streptomy

## ALIGNMENTS

RESULT 1  
ID ABB80941 standard; protein; 313 AA.

AC ABB80941;

DT 21-OCT-2002 (first entry)

DE S. coelicolor Scba protein.

KW Antibiotic, bacterium, scba, afra, scbr, arpa, bara, actinorhodine, Act;

KW undecylprodigiosin; Red.

OS Streptomyces coelicolor.

PN CA2322241-A1.

PD 23-APR-2002.

PF 23-OCT-2000; 2000CA-02322241.

PR 23-OCT-2000; 2000CA-02322241.

PA (PLAN-) PLANT BIOSCIENCE LTD.

PI Takano E, Bibb M;

DR WPI, 2002-501089/54.

XX N-PSDB; ABN6379.

PT Modifying antibiotic-producing Streptomyces, to increase, or alter timing

PT of, antibiotic production, by deleting the scba or scbr genes.

PS Claim 19; Fig 10; 64pp; English.

XX The invention provides a method for modifying an antibiotic-producing

XX strain of Streptomyces to increase production of antibiotics or to alter

XX the timing of antibiotic production. The modification is functional

XX deletion of the scba gene of S. coelicolor, or its homologues, but is not

XX functional deletion of the afra gene of S. griseus, or the modification is

XX deletion of the scbr gene of S. coelicolor, or its homologues, but is not

XX deletion of arpa of S. griseus nor bara of S. virginiae. The

XX method is particularly used for production of the antibiotics

XX actinorhodine (Act) and undecylprodigiosin (Red). The present sequence

XX represents the S. coelicolor Scba protein

XX Sequence 313 AA;

Query Match 100.0%; Score 1645; DB 5; Length 313;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-159; Mismatches 0; Indels 0; Gaps 0;  
 Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPEAVVLINSASDANSIEQTLAPVPMALVHRTRVQDAFPVSWIRPKGDRFSVTAVLPHDH 60  
 DB 1 MPEAVVLINSASDANSIEQTLAPVPMALVHRTRVQDAFPVSWIRPKGDRFSVTAVLPHDH 60  
 QY 61 PEFAPVHGDRHDPILLIAETLRQAAMLVFHAGYGVPGVGHFLMTLDYTCHELDHGVSGEVA 120  
 DB 61 PEFAPVHGDRHDPILLIAETLRQAAMLVFHAGYGVPGVGHFLMTLDYTCHELDHGVSGEVA 120  
 QY 121 ELEVEVACSQLKFRGGQPVQGVQVDMAVRAGRLAATGATTRFTSPVYRRMRGDFATPT 180  
 DB 121 ELEVEVACSQLKFRGGQPVQGVQVDMAVRAGRLAATGATTRFTSPVYRRMRGDFATPT 180  
 QY 181 ASVPGTAPVPAPARAGRTREDEVDVLSASSQODTWRLRVDTSHPTLFQRPNDHVPQMLLLEA 240  
 DB 181 ASVPGTAPVPAPARAGRTREDEVDVLSASSQODTWRLRVDTSHPTLFQRPNDHVPQMLLLEA 240  
 QY 241 AROAACLVTPGAPFPVPSIGTRFVRVYAEFDSPCWIOATVRPGPAAGLTTVRVTGHQDQSL 300  
 DB 241 AROAACLVTPGAPFPVPSIGTRFVRVYAEFDSPCWIOATVRPGPAAGLTTVRVTGHQDQSL 300  
 QY 301 VFLLTTLGGPAPFSG 313  
 DB 301 VFLLTTLGGPAPFSG 313

RESULT 2  
 ADEI4792  
 ID ADEI4792 standard; protein; 313 AA.  
 XX ADEI4792;  
 XX 29-JAN-2004 (first entry)  
 XX Streptomyces coelicolor ScdA protein.  
 DE Streptomyces coelicolor ScdA protein.  
 XX  
 KW antibiotic-producing strain; antibiotic production; scdA gene; ScdR gene;  
 KM actinorhodin; undecylprodigiosin.  
 XX  
 OS Streptomyces coelicolor.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 102  
 FT /note="Encoded by ATGGCC on the complementary strand of  
 FT the sequence in figure 14"  
 XX  
 XX US2003124644-A1.  
 XX  
 XX 03-JUL-2003.  
 XX  
 XX 23-OCT-2001; 2001US-00017471.  
 XX  
 XX 23-OCT-2000; 2000US-0242561P.  
 XX  
 XX 23-OCT-2000; 2000US-0242561P.  
 XX  
 XX (TAKA/) TAKANO E.  
 XX (BIBB/) BIBB M J.  
 XX  
 XX Takano E, Bibb MJ;  
 XX  
 XX WPI; 2003-810983/76.  
 XX N-PSDB; ADEI4790, ADEI4794.  
 XX  
 XX Modifying an antibiotic-producing strain of Streptomyces coelicolor or  
 XX Streptomyces lividans to increase or alter the timing of antibiotic  
 XX production in the strain, comprises functionally deleting in the strain  
 XX the scdA or ScdR gene.  
 XX  
 XX Claim 19; Fig 10; 33pp; English.  
 XX  
 XX This invention relates to the novel modification of an antibiotic-

CC producing strain of Streptomyces coelicolor or Streptomyces lividans to  
 CC increase or to alter the timing of antibiotic production in the strain.  
 CC The method comprises functionally deleting in the strain the scdA or ScdR  
 CC gene. The method is useful in increasing and altering the timing of  
 CC antibiotic production (especially actinorhodin and undecylprodigiosin) in  
 CC Streptomyces species, particularly Streptomyces coelicolor or  
 CC Streptomyces lividans. The present sequence is that of the Streptomyces  
 CC coelicolor ScdA protein which is used in the method of the invention.  
 CC  
 XX Sequence 313 AA;  
 XX SQ

Query Match 100.0%; Score 1645; DB 7; Length 313;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-159; Mismatches 0; Indels 0; Gaps 0;  
 Matches 313; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPEAVVLINSASDANSIEQTLAPVPMALVHRTRVQDAFPVSWIRPKGDRFSVTAVLPHDH 60  
 DB 1 MPEAVVLINSASDANSIEQTLAPVPMALVHRTRVQDAFPVSWIRPKGDRFSVTAVLPHDH 60  
 QY 61 PEFAPVHGDRHDPILLIAETLRQAAMLVFHAGYGVPGVGHFLMTLDYTCHELDHGVSGEVA 120  
 DB 61 PEFAPVHGDRHDPILLIAETLRQAAMLVFHAGYGVPGVGHFLMTLDYTCHELDHGVSGEVA 120  
 QY 121 ELEVEVACSQLKFRGGQPVQGVQVDMAVRAGRLAATGATTRFTSPVYRRMRGDFATPT 180  
 DB 121 ELEVEVACSQLKFRGGQPVQGVQVDMAVRAGRLAATGATTRFTSPVYRRMRGDFATPT 180  
 QY 181 ASVPGTAPVPAPARAGRTREDEVDVLSASSQODTWRLRVDTSHPTLFQRPNDHVPQMLLLEA 240  
 DB 181 ASVPGTAPVPAPARAGRTREDEVDVLSASSQODTWRLRVDTSHPTLFQRPNDHVPQMLLLEA 240  
 QY 241 AROAACLVTPGAPFPVPSIGTRFVRVYAEFDSPCWIOATVRPGPAAGLTTVRVTGHQDQSL 300  
 DB 241 AROAACLVTPGAPFPVPSIGTRFVRVYAEFDSPCWIOATVRPGPAAGLTTVRVTGHQDQSL 300  
 QY 301 VFLLTTLGGPAPFSG 313  
 DB 301 VFLLTTLGGPAPFSG 313

RESULT 3  
 ADH39754  
 ID ADH39754 standard; protein; 331 AA.  
 XX ADH39754;  
 XX 11-MAR-2004 (first entry)  
 XX Streptomyces rochei ORF85 protein SEQ ID NO:85.  
 DE Streptomyces rochei ORF85 protein SEQ ID NO:85.  
 XX  
 KW actinomycetes-originated plasmid; linear plasmid; pSLA2-L;  
 KM Streptomyces rochei; secondary metabolite; antibiotic; vitamin;  
 KW lankamycin; mithramycin-like substance; carotenoid; lankacidin.  
 XX  
 OS Streptomyces rochei.  
 XX  
 PN WO2004001039-A1.  
 XX  
 XX 31-DEC-2003.  
 XX  
 XX 19-JUN-2003; 2003WO-JP007767.  
 XX  
 XX 20-JUN-2002; 2002JP-00179345.  
 XX  
 XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
 XX  
 XX Kinashi H;  
 XX  
 XX WPI; 2004-082503/08.  
 XX  
 XX Actinomycetes-originated plasmids for producing secondary metabolites e.g.  
 XX antibiotics and vitamins like lankamycins, lankacidins, mithramycin-like  
 XX substances and carotenoids.



XX	Claim 1; SEQ ID NO 85; 499pp; Japanese.
XX	
XX	The present invention describes an actinomycetes-originated plasmid is a
CC	linear plasmid (pSLA2-L) originating in Streptomyces rochei, which has a
CC	base sequence capable of providing any of the protein sequences of SEQ ID
CC	NO.1 to 143 (ADH39670 to ADH39812), or one based on these sequences but
CC	with some amino acids deleted, substituted or added and capable of
CC	producing a secondary metabolite. Also described is a process for
CC	producing a secondary metabolite by using the plasmid. The plasmid can be
CC	used for producing secondary metabolites e.g. antibiotics and vitamins
CC	like lankmycins, mithramycin-like substances, carotenoids and
CC	lanketadins. Such secondary metabolites are produced selectively and
CC	efficiently. The present sequence represents a Streptomyces rochei
CC	protein.
XX	
XX	Sequence 331 AA;
XX	
QY	Query Match 31.9%; Score 524.5; DB 8; Length 331;
Db	Beet Local Similarity 38.4%; Pred. No. 1,2e-44;
	Matches 118; Conservative 44; Mismatches 136; Indels 9; Gaps 5
QY	10 SASDANSIEGTALPPPMALVHRTVRQDAFAPSWIKSGDRFSVTAVLPHDHFAPVAGD 69
Db	23 SPGDTSMWLPITLTVPRREYVIRASLAERFLTRCRTHETRFLLTCQWRRAHTFFLSPPGR 82
QY	70 RHDPDLIAETLRQAAMLVFHAGYGVVGYHFLM-TLDYTCILDHLGVSGEVAELEEVAC 128
Db	83.RHDPQIAETLRQVGLHLAAHAFVDVPLGHFFIMDMDSFVSRVHGLGVGRPTDLDVEATC 142
QY	129 SOLKFRGGQAPVQGVDAVVRAGRLAATGTATRT--SPQYRRMRGDFATPTASVGT 186
Db	143 VDVAARRRCKLVDFRVLVITIEDGHLVANGG--RFTCTTEAMRYRLRSAPATTAAQAS 200
QY	187 ---APVPARARGTDEDEVVLSASSQDQTLRLRVDTSHPTLFQRNDHPVPGALLLEAARQ 243
Db	201 HQPAPLPPSDRGRTAPRDVLAPEGAPKMRRLNADTSHLILDHGCDVPGWVLESARQ 260
QY	244 AACTLVGP-APFVPSIGSTRFVRVAFEDSPCWIQATVBPGRPAAGLITTVRGHODGSIVF 302
Db	261 AACAHLPPGSLIPRTVSTEFRRVYEFSPCIEASGLAVTSGTFHALLITGRDDDEVF 320
QY	303 LTTLSGP 309
Db	321 TARISGP 327
XX	
XX	RESULT 4
ID	ABB80938
XX	ABB80938 standard; protein; 57 AA.
AC	ABB80938;
XX	
DT	21-OCT-2002 (first entry)
XX	
DE	S. coelicolor scda protein fragment.
KW	Antibiotic; bacterium; scda; afsA; scdR; arpA; barA; actinorhodine; Act;
KM	undecylprodigiosin; Red.
XX	
OS	Streptomyces coelicolor.
PN	CA322241-A1.
XX	
XX	23-APR-2002.
PF	23-OCT-2000; 2000CA-02322241.
XX	
PR	23-OCT-2000; 2000CA-02322241.
XX	
PA	(PLAN-) PLANT BIOSCIENCE LTD.
XX	
PI	Takano E, Bibb M;

XX	WPI: 2002-501089/54.
DR	N-PSDB; AEN86377.
XX	
PT	Modifying antibiotic-producing Streptomycetes, to increase, or alter timing
PT	of, antibiotic production, by deleting the scbA or scbR genes.
XX	
PS	Disclosure; Fig 4B; 64pp; English.
XX	
CC	The invention provides a method for modifying an antibiotic-producing
CC	strain of Streptomycetes to increase production of antibiotics or to alter
CC	the timing of antibiotic production. The modification is functional
CC	deletion of the scbA gene of S. coelicolor, or its homologues, but is not
CC	deletion of the afbA gene of S. griseus, or the modification is
CC	functional deletion of the scbR gene of S. coelicolor, or its homologues,
CC	but is not deletion of atpA of S. griseus nor baba of S. viridulae. The
CC	method is particularly used for production of the antibiotics
CC	actinorhodine (Act) and undecylprodigiosin (Red). The present sequence
XX	represents the S. coelicolor ScbA protein fragment
XX	
SQ	Sequence 57 AA;
XX	
Query Match	17.5%; Score 288; DB 5; Length 57;
Best Local Similarity	100.0%; Pred. No. 1.5e-21;
Matches	57; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Oy	1 MPEAVVLINSASDANSIEQTALPVPMAVLVHRTVODAFPVSWIPKGDREFTAVLP 57
Db	1 MPEAVVLINSASDANSIEQTALPVPMAVLVHRTVODAFPVSWIPKGDREFTAVLP 57
RESULT 5	
AAU04042	
ID	AAU04042 standard; protein; 353 AA.
XX	
AC	AAU04042;
XX	
DT	11-SEP-2003 (revised)
DT	23-OCT-2001 (first entry)
XX	
DE	Streptococcus coelicolor Mmfl protein.
XX	
SCPI	SCP1; methylenomycin cluster; mmc; MmyR; MmfP; MmfH; MmfR; MmyT;
KM	MmyO; MmyG; MmyJ; Mmr; heterologous gene expression.
XX	
OS	Streptomycetes coelicolor.
XX	
PN	WO200148228-A1.
XX	
PD	05-JUL-2001.
XX	
PF	20-DEC-2000; 2000MO-GB004972.
XX	
PR	23-DEC-1999; 99GB-00030477.
XX	
PA	(PLAN-) PLANT BIOSCIENCE LTD.
XX	
PI	Chater KF, Bruton CJ, O'Rourke SJ, Wietzorrek AW;
XX	
DR	WPI: 2001-425675/45.
DR	N-PSDB; AAS07627; AAU04045, AAU04046.
XX	
PT	Novel expression cassette for expressing a nucleic acid of interest,
PT	derived from the regulatory region of methylenomycin gene cluster of SCP1
XX	plasmid of Streptomycetes coelicolor A3(2).
XX	
PS	Claim 25; Fig 8d; 142pp; English.
XX	
CC	The sequence represents the Mmfl protein encoded by the mmfl gene carried
CC	on the expression cassette present on plasmid SCP1. The expression
CC	cassette is the regulatory region of the methylenomycin cluster (mmc)
CC	from Streptomycetes coelicolor A3(2), which encodes the MmyR, MmfP, MmfH,
CC	MmfL, MmyT, MmyG, MmyJ and partial Mmr polypeptides. The

CC expression cassette is useful for expressing a nucleic acid of interest,  
 CC substantially only when the host cell culture reaches high cell density  
 CC at or close to the stationary phase of host cell culture. In particular  
 CC the system is useful in regulating methylenomycin production. Reduced or  
 CC no expression of the nucleic acid of interest is observed earlier in  
 CC growth, avoiding toxic effects of some gene products on growth and the  
 CC system does not require addition of exogenous inducer. The methylenomycin  
 CC cluster naturally present on a highly transmissible plasmid permits  
 CC properly regulated expression in diverse Streptomyces host and the  
 CC expression is driven by a strong promoter, leading to high yield of the  
 CC desired end product. (Updated on 11-SEP-2003 to standardise OS field)

XX  
 SQ Sequence 353 AA;

Query Match 16.5%; Score 271.5; DB 4; Length 353;  
 Best Local Similarity 27.9%; Pred. No. 1e-18;  
 Matches 95; Conservative 43; Mismatches 123; Indels 79; Gaps 12;

QY 21 ALPVMALVHRTVODAFVSVIPKGGDFSTAVLPHDHPFAVHGDRHPLLAETL 80  
 DB 26 AAPLPADVHKAAMAEVLLTDARPLGENRFAVAALMPRTFLAHRATSSPCDPLAETI 85  
 QY 81 RQAMLVFAGYGVPGVFLMTLDYTHLDHAGVGEVALEVEVACSGQLKFRGGQPVQ 140  
 DB 86 RQSAHLHSTFCDDVPIGHFVL-----SGLDLDLDLVWDS-----GPLEVY 127  
 QY 141 GQVMAVRBAGLAAATGATTT-----RFT-TSPQVY--- 169  
 DB 128 -----LDVTSTKTTNPRMARALNADYVAGLRGCAIRFVLAAPRYAMI 175  
 QY 170 --RRRRGDPATPTASVPGTAPPARAGTRDEVDVL-SASSQOFT-WLRVDTHPTLF 225  
 DB 176 RDRARRAEPPAOQAAAGATATLPEPTVGHDDHLVLAHQGLPTAQQLRRDHPVLF 235  
 QY 226 QRPNDHVPGLTLEAARQAACVTGPAP--FVP-----SIGTRFVRYAEPDPCMIQAT 278  
 DB 236 DHESHISMALLECROAATLTTPAPCAFSPROVALTAVASSVOAREGLDSP--VIT 293  
 QY 279 VAP-----GPAAGLTVTVYGHQDGSVLPVLTTLGPAFSG 313  
 DB 294 TLPAAHGHPDSDGTRTLQTLARQGSRTLTITAVTTTGTG 333

RESULT 6  
 ADCC31753  
 ID ADCC31753 standard; protein; 541 AA.

AC ADCC31753;

DT 18-DEC-2003 (first entry)

DE Human novel polypeptide sequence, SEQ ID NO.1835.

XX Human; diagnostic; drug screening; forensics; gene mapping;  
 KM biodiversity assessment; Parkinson's disease; Alzheimer's disease;  
 KM neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
 KM ulcers; osteoporosis; autoimmune disease; cancer;  
 KM molecular weight marker; food supplement; antiparkinsonian; nootropic;  
 KM neuroprotective; anti-anemic; anticoagulant; thrombolytic; vulnerary;  
 KM antifungal; osteopathic; immunosuppressive; antiinflammatory; cycostatic;  
 KM gene therapy; chromosome 3.

XX Homo sapiens.

XX WO2003029271-A2.

PD 10-APR-2003.

XX 24-SEP-2002; 2002WO-US030474.

XX 24-SEP-2001; 2001US-0324631P.

XX (HYSE-) HYSEQ INC.

XX Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;  
 PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
 PI Haley-Vicente D, Dmanac RT;  
 DR MPI; 2003-371981/35.  
 DR N-PSDB; ADCC30782.

XX New polynucleotide and polypeptide useful for diagnosing, preventing or  
 PT treating conditions such as neurodegenerative diseases, anemias, platelet  
 PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
 PT cancer.

PS Claim 20; SEQ ID NO 1835; 1185bp; English.

XX The invention relates to 971 novel human cDNA sequences (ADC29919-  
 CC ADCC30889) and the polypeptides they encode (ADCC30890-ADCC31860). The  
 CC invention also relates to nucleic acid sequences over 99% identical with  
 CC the novel human cDNAs. The invention additionally encompasses expression  
 CC vectors and host cells comprising a nucleic acid of the invention; the  
 CC recombinant production of a polypeptide of the invention; an antibody  
 CC against a polypeptide of the invention; a method of detecting  
 CC polynucleotides or polypeptides of the invention; and methods of  
 CC identifying a compound which binds to a polypeptide of the invention. The  
 CC invention further discloses methods of preventing, treating or  
 CC ameliorating a medical condition; kits comprising polynucleotide probes  
 CC and/or monoclonal antibodies for carrying out the methods of the  
 CC invention; methods for the identification of compounds that modulate the  
 CC expression or activity of the polynucleotide and/or polypeptide; and 767  
 CC coding sequences corresponding to the cDNA sequences of the invention  
 CC (ADCC31861-ADCC32627) and the polypeptides encoded by the cDNAs (ADCC32628  
 CC -ADCC3394). The nucleic acids and polypeptides of the invention are  
 CC useful in diagnostics, drug screening, forensics, gene mapping, in the  
 CC identification of mutations responsible for genetic disorders or other  
 CC traits, for assessing biodiversity, and in producing many other types of  
 CC data and products dependent on DNA and amino acid sequences. They are  
 CC also used for treating diseases such as Parkinson's disease, Alzheimer's  
 CC disease and other neurodegenerative diseases, anaemia, platelet  
 CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
 CC cancer. The nucleic acids may also be used as hybridisation probes or  
 CC primers, and in the recombinant production of a protein. The polypeptides  
 CC are also useful in generating antibodies, as molecular weight markers,  
 CC and as food supplements. The present sequence represents a specifically  
 CC claimed human polypeptide sequence of the invention. Note: The sequence  
 CC data for this patent did not form part of the printed specification, but  
 CC was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 541 AA;

Query Match 7.7%; Score 126; DB 7; Length 541;  
 Best Local Similarity 28.6%; Pred. No. 0.0014;  
 Matches 55; Conservative 21; Mismatches 30; Indels 26; Gaps 7;

QY 50 FSVAVLPHDHPFAPVHGDRHDL-----LAEITRQAAMLVFNHGVGVPGYHFLM 102

DB 77 FSLAL-----HPFQVTHREKPTPEHIFRAILIKKIDRSQKTDLHNGY-----IL 124

QY 103 TLDTYCHLDHGVSGEVAEL--EVEVACSG--LKFRGGQVPGVDMVRRAGRLAATGTA 159

DB 125 ELDCSSLDHPTDQKLTPEFTKIQEAAASGQLKRVGVF---QTHSSVNSAGSAPVSTA 181

QY 160 TTRFTSPQVYRRMGDFATPTASVGTAPVAPARAGTRDEVDVLSSAQDTRRLAVDT 219

DB 182 NSTEDARDA-KAAGDHASLENEKGTDCVAPAGRQSPSPSGPRGVEHEEPLSGKM 240

QY 220 SHPTLFQRPNDH 231

DB 241 EIFTLFNPKSH 252

RESULT 7  
 ADCC31754

ADG31754 standard; protein; 516 AA.  
 ADG31754;  
 18-DEC-2003 (first entry)  
 Human novel polypeptide sequence, SEQ ID NO:1836.  
 Human; diagnostic; drug screening; forensics; gene mapping;  
 biodiversity assessment; Parkinson's disease; Alzheimer's disease;  
 neurodegenerative diseases; anaemia; platelet disorder; wound; burns;  
 ulcers; osteoporosis; autoimmune disease; cancer;  
 molecular weight marker; food supplement; antiparkinsonian; nootropic;  
 neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnary;  
 antifuror; osteopathic; immunosuppressive; antiinflammatory; cyostatic;  
 gene therapy; chromosome 3.  
 Homo sapiens.  
 WO2003029271-A2.  
 10-APR-2003.  
 24-SEP-2002; 2002WO-US030474.  
 24-SEP-2001; 2001US-0324631P.  
 (HYSE-) HYSEQ INC.  
 Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehman T;  
 Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;  
 Haley-Vicente D, Drmanac RT;  
 WPI; 2003-371981/35.  
 N-PSDB; ADG30783.  
 New polynucleotide and polypeptide useful for diagnosing, preventing or  
 treating conditions such as neurodegenerative diseases, anemias, platelet  
 disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
 cancer.  
 Claim 20; SEQ ID NO 1836; 1185bp; English.  
 The invention relates to 971 novel human cDNA sequences (ADG29919-  
 ADG30889) and the polypeptides they encode (ADG30890-ADG31860). The  
 invention also relates to nucleic acid sequences over 99% identical with  
 the novel human cDNAs. The invention additionally encompasses expression  
 vectors and host cells comprising a nucleic acid of the invention; the  
 recombinant production of a polypeptide of the invention; an antibody  
 against a polypeptide of the invention; a method of detecting  
 polynucleotides or polypeptides of the invention; and methods of  
 identifying a compound which binds to a polypeptide of the invention. The  
 invention further discloses methods of preventing, treating or  
 ameliorating a medical condition; kits comprising polynucleotide probes  
 and/or monoclonal antibodies for carrying out the methods of the  
 invention; methods for the identification of compounds that modulate the  
 expression or activity of the polynucleotide and/or polypeptide; and 767  
 coding sequences corresponding to the cDNA sequences of the invention  
 (ADG31861-ADG32627) and the polypeptides encoded by the contigs (ADG32628  
 -ADG33394). The nucleic acids and polypeptides of the invention are  
 useful in diagnostics, drug screening, forensics, gene mapping, in the  
 identification of mutations responsible for genetic disorders or other  
 traits, for assessing biodiversity, and in producing many other types of  
 data and products dependent on DNA and amino acid sequences. They are  
 also used for treating diseases such as Parkinson's disease, Alzheimer's  
 disease and other neurodegenerative diseases, anaemia, platelet  
 disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
 cancer. The nucleic acids may also be used as hybridisation probes or  
 primers, and in the recombinant production of a protein. The polypeptides  
 are also useful in generating antibodies, as molecular weight markers,  
 and as food supplements. The present sequence represents a specifically  
 claimed human polypeptide sequence of the invention. Note: The sequence  
 data for this patent did not form part of the printed specification, but

was obtained in electronic format directly from WIP0 at  
 ftp.wipo.int/pub/published\_pct\_sequences.  
 Sequence 516 AA;  
 Query Match 7.2%; Score 118.5; DB 7; Length 516;  
 Best Local Similarity 27.1%; Pred. No. 0.0077;  
 Matches 62; Conservative 22; Mismatches 82; Indels 63; Gaps 11;  
 50 FSVYAVLPHPHPFAPVPGDRHDP-----LIAETLRQAMLVFAGYGVPGVYHFLM 102  
 77 FSLAAL-----HPFQPHREKPTLPHIIFRAILLIKTRDROKQDLNHEG-----IL 124  
 103 TLDYTCCHDLGVSGEVAEL--EVEVACSQ-LKFRGQPVQGVQVDAVRAGRILATGTA 159  
 125 ELDCSSLDHPDQKILPEFTIKKIQAASQGLKFGVYIP---QVHSSVNSAGSSAPVSTA 181  
 160 TTRFTSPQVTRRRMGDPTPTASVPGTAPYPAARAAGTRDE-----DVL-----SA 206  
 182 NSTEDARDA-KNARGDHASDENEKPTGVDYCSAPAGNQSPEPSSGPRGEVPLAKQSSP 240  
 207 SSGQDTRLRVD-----TSHP-----TLFORPNDH 231  
 241 SGEQDGEELSPQGVSKTLDGPESNPLEVHEEPLSGKMEITTLPRKPSH 289  
 RESULT 8  
 ABM80427  
 ID ABM80427 standard; protein; 578 AA.  
 AC ABM80427;  
 XX 18-NOV-2004 (first entry)  
 XX Tumour-associated antigenic target (TAT) polypeptide PRO80962, SEQ:1073.  
 DE Tumour-associated antigenic target; TAT; human; overexpression; cancer;  
 KW tumour; diagnosis; cell proliferative disorder; breast cancer;  
 KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;  
 KW central nervous system cancer; bladder cancer; pancreatic cancer;  
 KW cervical cancer; melanoma; leukaemia; hybridisation probe;  
 KW chromosome identification; chromosome mapping; gene mapping;  
 KW gene therapy; cyostatic.  
 OS Homo sapiens.  
 XX WO2004030615-A2.  
 XX 15-APR-2004.  
 XX 29-SEP-2003; 2003WO-US028547.  
 XX 02-OCT-2002; 2002US-0414971P.  
 XX (GENTH ) GENENTECH INC.  
 XX Wu TD, Zhang Z, Zhou Y;  
 PI WPI; 2004-347921/32.  
 DR N-PSDB; ACN37889.  
 XX New tumor-associated antigenic target polypeptides and nucleic acids,  
 PT useful in preparing a medicament for treating or detecting a  
 PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or  
 PT prostate cancer or tumor.  
 XX Claim 12; SEQ ID NO 1073; 7273bp; English.  
 XX The invention relates to human tumour-associated antigenic target (TAT)  
 CC polypeptides, and their related nucleic acids. The TAT polypeptides are  
 CC overexpressed in cancer tissues compared to normal tissues, and may thus  
 CC serve as effective targets for the diagnosis and treatment of cancer in  
 CC mammals. The invention also relates to nucleic acid and polypeptide

CC sequences at least 80% identical to the TAT nucleic acids and  
 CC polypeptides; expression vectors and host cells comprising a TAT nucleic  
 CC acid; an antibody specific for a TAT polypeptide; a peptide or organic  
 CC molecule which binds to a TAT polypeptide; fusion proteins comprising a  
 CC TAT polypeptide; and methods and compositions for the treatment or  
 CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,  
 CC antibodies, antagonists, binding molecules and compositions are useful  
 CC for diagnosing or treating a cell proliferative disorder associated with  
 CC increased TAT expression, particularly cancers such as breast cancer,  
 CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder  
 CC cancer, pancreatic cancer, cervical cancer, cancers of the central  
 CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be  
 CC used as hybridisation probes, in chromosome and gene mapping, in  
 CC chromosome identification and in gene therapy. The present sequence  
 CC represents a TAT polypeptide of the invention

CC Sequence 578 AA;

Query Match 7.2%; Score 118.5; DB 8; Length 578;  
 Best Local Similarity 27.1%; Pred. No. 0.009; Mismatches 82; Indels 63; Gaps 11;  
 Matches 62; Conservative 22; Mismatches 82; Indels 63; Gaps 11;

QY 50 FSVTVLPHDHPFAPVGHDRDPL-----LIAETLRQAAMLVFHAGYGVGVYHFLM 102  
 DB 77 FSLAAL---HPVQPTHERKPTLEHIFRALIKKIDRSQKTDLHNGY-----IL 124  
 QY 103 TLDTYCHDLHGVSEVAEL--EVEVACSQ--LKFPGGQPVGGQVMAVRARGLAATGTA 159  
 DB 125 ELDCSSLDHPTDQKLIPEFIKKIQEASQGLKFGVIP---QYHSSVNASGSSAPVSTA 181  
 QY 160 TTRFTSPQVYRRMGDFATPTASVGTAPVPAARAGRTRDE-----DVL-----SA 206  
 DB 182 NSTEDARDA-KNARGDHASLENEKPGTGVCSAPAGRQSPSPSSGPRGEVPLAKQPSSP 240  
 QY 207 SSGQDPTWLRVD-----TSHP-----TLFORPNHD 231  
 DB 241 SSGGDSGLSPQGVSKTLDGPESNPLFVHEERPLSGKMEITFLFKPKSH 289

RESULT 9  
 ADR14454  
 ID ADR14454 standard; protein; 648 AA.

AC ADR14454;

DT 21-OCT-2004 (first entry)

DE Human NF-kappaB pathway-associated protein Seg1D455.

KW NF-kappaB pathway; antiinflammatory; cytostatic; hepatotropic; virucide;  
 KW antiarthritic; antirheumatic; gastrointestinal-gen; antiallthmatic;  
 KW antiarteriosclerotic; immunomodulator; cerebroprotective; vasotrophic;  
 KW immunosuppressive; vulnerary; gene therapy; immune disorder;  
 KW inflammatory disorder; NF-kappaB regulation; cancer; aberrant apoptosis;  
 KW hepatic disorder; Hodgkin's lymphoma; haematopoietic tumour;  
 KW hyper-IgM syndrome; hypohidrotic ectodermal dysplasia;  
 KW X-linked ahidrotic ectodermal dysplasia; immunodeficiency;  
 KW viral infection; HIV-1; HTLV-1; hepatitis B; hepatitis C; EBV; influenza;  
 KW viral replication; host cell survival; evasion of immune response;  
 KW rheumatoid arthritis; inflammatory bowel disease; colitis; asthma;  
 KW atherosclerosis; cachexia; euthyroid sick syndrome; stroke; EAE;  
 KW autoimmune acute phase response; hyper immune activity;  
 KW aberrant acute phase response; hypercongenital condition; birth defect;  
 KW necrotic lesion; wound; organ transplant rejection;  
 KW aberrant signal transduction; proliferating disorder; cancer;  
 KW HIV propagation; human.

XX Homo sapiens.

OS WO200406577-A2.

PN 05-AUG-2004.

XX

PF 13-JAN-2004; 2004WO-US000798.

XX 14-JAN-2003; 2003US-0440068P.

PR 12-MAY-2003; 2003US-0469757P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

PI Nadler SG, Neubauer MG, Feder JN, Carman J;

XX WPI; 2004-562168/54.

DR N-PSDB; ADR14455.

PT New isolated polynucleotides and polypeptides associated with NF-kappaB  
 PT pathway, useful for diagnosing, treating, or preventing disorders or  
 PT diseases associated with NF-kappaB pathway.

PS Claim 1; SEQ ID NO 455; 237pp; English.

XX This invention relates to the novel association of protein sequences (and  
 CC the genes which encode them) to the NF-kappaB pathway. The invention may  
 CC be useful for the production of compounds with an antiinflammatory,  
 CC cytostatic, hepatotropic, virucide, antirheumatic, antineumatic,  
 CC gastrointestinal-gen, antialthmatic, antiarteriosclerotic,  
 CC immunomodulator, cerebroprotective, vasotropic, immunosuppressive or  
 CC vulnerary activity or for gene therapy. The proteins and nucleotides are  
 CC useful for diagnosing, preventing, treating, or ameliorating conditions  
 CC or diseases associated with the NF-kappaB pathway. The condition is an  
 CC immune disorder, an inflammatory disorder, an inflammatory disorder  
 CC related to aberrant NF-kappaB regulation, cancer, aberrant apoptosis,  
 CC hepatic disorders, Hodgkin's lymphomas, haematopoietic tumours, hyper-IgM  
 CC syndromes, hypohidrotic ectodermal dysplasia, X-linked ahidrotic  
 CC ectodermal dysplasia, immunodeficiency, viral infections, HIV-1, HTLV-1,  
 CC hepatitis B, hepatitis C, EBV, influenza, viral replication, host cell  
 CC survival, evasion of immune responses, rheumatoid arthritis, inflammatory  
 CC bowel disease, colitis, asthma, atherosclerosis, cachexia, euthyroid sick  
 CC syndrome, stroke, EAE, autoimmune disorders, disorders related to hyper  
 CC immune activity, disorders related to aberrant acute phase responses,  
 CC hypercongenital conditions, birth defects, necrotic lesions, wounds,  
 CC organ transplant rejection, conditions related to organ transplant  
 CC rejection, disorders related to aberrant signal transduction,  
 CC proliferating disorders, cancers and HIV propagation in cells infected  
 CC with other viruses. The present sequence is that of a human protein which  
 CC is subject to the novel association with the NF-kappaB pathway of the  
 CC invention. Note: This sequence does not appear in the specification but  
 CC was obtained by the indexer from Genbank.

XX Sequence 648 AA;

Query Match 7.2%; Score 118.5; DB 8; Length 648;  
 Best Local Similarity 27.1%; Pred. No. 0.011; Mismatches 82; Indels 63; Gaps 11;  
 Matches 62; Conservative 22; Mismatches 82; Indels 63; Gaps 11;

QY 50 FSVTVLPHDHPFAPVGHDRDPL-----LIAETLRQAAMLVFHAGYGVGVYHFLM 102  
 DB 147 FSLAAL---HPVQPTHERKPTLEHIFRALIKKIDRSQKTDLHNGY-----IL 194  
 QY 103 TLDTYCHDLHGVSEVAEL--EVEVACSQ--LKFPGGQPVGGQVMAVRARGLAATGTA 159  
 DB 195 ELDCSSLDHPTDQKLIPEFIKKIQEASQGLKFGVIP---QYHSSVNASGSSAPVSTA 251  
 QY 160 TTRFTSPQVYRRMGDFATPTASVGTAPVPAARAGRTRDE-----DVL-----SA 206  
 DB 252 NSTEDARDA-KNARGDHASLENEKPGTGVCSAPAGRQSPSPSSGPRGEVPLAKQPSSP 310  
 QY 207 SSGQDPTWLRVD-----TSHP-----TLFORPNHD 231  
 DB 311 SSGGDSGLSPQGVSKTLDGPESNPLFVHEERPLSGKMEITFLFKPKSH 359

RESULT 10  
 ABO72006  
 ID ABO72006 standard; protein; 749 AA.

XX

AC ABO72006;  
XX 29-JUL-2004 (first entry)  
XX Pseudomonas aeruginosa polypeptide #4181.  
XX Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.  
XX Pseudomonas aeruginosa.  
XX US6551795-B1.  
XX 22-APR-2003.  
XX 18-FEB-1999; 99US-00252991.  
XX 18-FEB-1998; 98US-0074788P.  
XX 27-JUL-1998; 98US-0094190P.  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX WPI; 2003-615309/58.  
XX N-PSDB; ABD05577.  
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
XX PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
XX patho logical conditions resulting from bacterial infection.  
XX Disclosure; SEQ ID NO 20752; 455pp; English.  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
XX polynucleotides encoding them. The sequences are useful in diagnosis and  
XX therapy of pathological conditions, as molecular targets for diagnostics,  
XX prophylaxis and treatment of pathological conditions resulting from a  
XX bacterial infection, for evaluating a compound, such as a polypeptide,  
XX for the ability to bind a P. aeruginosa nucleic acid, as components of  
XX effective antibacterial targets, as targets for antibacterial drugs,  
XX including anti-P. aeruginosa drugs, as templates for recombinant  
XX production of P. aeruginosa-derived peptides or polypeptides, as target  
XX components for diagnosis and/or treatment of P. aeruginosa-caused  
XX infection, and in detection of P. aeruginosa sequences or other sequences  
XX of Pseudomonas species using biochip technology. Sequences ABO67826-  
XX ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The  
XX sequence data for this patent did not form part of the printed  
XX specification but was obtained in electronic format from USPTO at  
XX seqdata.uspto.gov/sequence.html  
SQ Sequence 749 AA;  
Query Match 6.7%; Score 110; DB 7; Length 749;  
Best Local Similarity 24.7%; Pred. No. 0.097;  
Matches 74; Conservative 26; Mismatches 116; Indels 84; Gaps 15;  
43 IPKGGDRSVTAVALPHDHPFAPVYGDHDDLLIA-----ETLRQAAMLVFVHAGYVVP 95  
DB 465 VPEREDLAGMNVLAGHGLADAGEGHDAIVAGFQQLLDITGDLPGHLPFGAG-ARP 523  
QY 96 VGYHFLMTLDYTCCHDLGVSGEVALEEV-EVACSQLKFRGGQPVQGVDMVAV--RRAR 152  
DB 524 VGLH-----HHGLSEGERVFPALPOLAVGQRADRAKQDHQVDDLAIIQRPQGX 571  
QY 153 LAATGTATT-----RTSPQVYRRMRGDFATPTASVPG--TAVPARAGATRR 198  
DB 572 V--EGHATTMPWPVSPCGMPGRSDAPALREAPGRMPRPPSGRLAPAPGP--NHR 626  
QY 199 DEDVVLASSQODTRLRVTLSHPTLFGRPNDHVGMLLEAARQAACLVNTP-----251  
DB 627 DRCAPPAAG-----SSCPA--RPPARTTGRPLRSHPTREP--ITAPRRRRRR 671  
QY 252 ----APFVPSIGSTRFV-----RYAEFDS-----CWIQATVRPQPAAGLTIVR 291

DB 672 PARSGPVPRSGRRRPPGRRRYAIDRRPARARAGVARRSVCVATRVPTPAGRCGNVR 731  
RESULT 11  
ID ABO84125  
ID ABO84125 standard; protein; 660 AA.  
XX ABO84125;  
XX 29-JUL-2004 (first entry)  
XX Pseudomonas aeruginosa polypeptide #16300.  
XX Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.  
XX Pseudomonas aeruginosa.  
XX US6551795-B1.  
XX 22-APR-2003.  
XX 18-FEB-1999; 99US-00252991.  
XX 18-FEB-1998; 98US-0074788P.  
XX 27-JUL-1998; 98US-0094190P.  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX WPI; 2003-615309/58.  
XX N-PSDB; ABD17696.  
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
XX PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
XX patho logical conditions resulting from bacterial infection.  
XX Disclosure; SEQ ID NO 32871; 455pp; English.  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
XX polynucleotides encoding them. The sequences are useful in diagnosis and  
XX therapy of pathological conditions, as molecular targets for diagnostics,  
XX prophylaxis and treatment of pathological conditions resulting from a  
XX bacterial infection, for evaluating a compound, such as a polypeptide,  
XX for the ability to bind a P. aeruginosa nucleic acid, as components of  
XX effective antibacterial targets, as targets for antibacterial drugs,  
XX including anti-P. aeruginosa drugs, as templates for recombinant  
XX production of P. aeruginosa-derived peptides or polypeptides, as target  
XX components for diagnosis and/or treatment of P. aeruginosa-caused  
XX infection, and in detection of P. aeruginosa sequences or other sequences  
XX of Pseudomonas species using biochip technology. Sequences ABO67826-  
XX ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The  
XX sequence data for this patent did not form part of the printed  
XX specification but was obtained in electronic format from USPTO at  
XX seqdata.uspto.gov/sequence.html  
SQ Sequence 660 AA;  
Query Match 6.3%; Score 104; DB 7; Length 660;  
Best Local Similarity 25.9%; Pred. No. 0.33;  
Matches 72; Conservative 17; Mismatches 131; Indels 58; Gaps 12;  
34 VQDAFPVSWIPKGGDRSVTAVALPHDHPFAPVYGDHDDLLIAETLRQAAMLVFVHAGY 92  
DB 1 IPSAMPSPSRPCCMPR-----RPSDDP-----GVN-----PLQILPLRRPARLHANRPGM 46  
QY 93 GVPVGYHFLMTLDY--TCHLDLGVSGEVALEEVACSQLKFRGGQPVQGVDMVAV--RRA 150  
DB 47 SAARARHRLYADIGGRARGLRSAGRGALDIOAPRAQREVNHVGPRRPQADAPGADRL 106  
QY 151 GRILATGTATRTFTSPQVYRRMRGDFATPTASVPGTAVPADAGRTREDVVLASSSQ 210  
DB 107 QODARTGCLLR--RRLFQKGPDLAOP--AVPVPGLPLR-----144



```

Oy 114 GVSQGEVLEVEVACSOQK-----FRGGQPVGGQVDMWVRBAGRLAATGTAT 160
Db 295 -VHPRALLPAVAVNCDDDKTGSGSEANGRTLARELGNTSFGQ-----RTGAVWVTGSSQ 346
Oy 161 TRFTSP-QVYRRMGDFATPTASVPGTAP-----VPARAGRTREDDVLSASSQDPTW 213
Db 347 VHVSRHFRVWGLSRDPLGAPAMATVGSWQDQGLDFQGAALR-----VPSPGTQARP 400
Oy 214 RLRYDT--SHPTLFQRPNDH---VPGMLL-----EAAR---QAACLVTPGAP 253
Db 401 KLRVVTLVHPFVFTRESDEGQCPAGQLCLDPTGNTSARLDALFALVNGSV 454

```

## RESULT 14

```

ADCT72229
ID ADCT72229 standard; protein; 987 AA.
XX
XX AC ADCT72229;
XX
XX 18-DEC-2003 (first entry)
XX
XX DE Rat NR3B SEQ ID NO:2.
XX
XX rat; NR3B: cerebroprotective; nootropic; antiparkinsonian; gene therapy;
XX stroke; Alzheimer's disease; Parkinson's disease; N-methyl-D-aspartate;
XX NMDA; N-methyl-D-aspartate receptor type 3B.
XX
XX OS Rattus sp.
XX
XX PN WO2003016479-A2.
XX
XX PD 27-FEB-2003.
XX
XX PF 16-AUG-2002; 2002WO-US026022.
XX
XX PR 20-AUG-2001; 2001US-00934070.
XX
XX PA (BURN-) BURNHAM INST.
XX PA (UYVA ) UNIV YALE.
XX PA (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX
XX PI Lipton SA, Zhang D, Chatterton JE, Awobuluyi M, Sevarino KA;
XX
XX DR WPI; 2003-278559/27.
XX
XX PT N-PSDB; ADCT72228.
XX
XX PT New NR3B nucleic acid molecule encoding a NR3B polypeptide, useful for
XX preparing a composition for treating e.g. stroke, Alzheimer's disease or
XX Parkinson's disease.
XX
XX PS Claim 21; SEQ ID NO 2; 190pp; English.
XX
XX CC The invention relates to a novel NR3B nucleic acid molecule. A NR3B
XX protein has cerebroprotective, nootropic, and antiparkinsonian activity.
XX A gene of the invention may have a use in gene therapy. The NR3B nucleic
XX acid is useful for preparing a composition for treating e.g. stroke,
XX Alzheimer's disease or Parkinson's disease. The present sequence is used
XX in the exemplification of the invention.
XX
XX SQ Sequence 987 AA;

```

```

Query Match 6.2%; Score 101.5; DB 7; Length 987;
Best Local Similarity 23.8%; Pred. No. 1.1;
Matches 70; Conservative 31; Mismatches 96; Indels 97; Gaps 13;

```

```

Oy 1 MPEAVVLINSASDANSIEQTALPVPMALVHRTVQDAFPVSWIPKG-----GDRFSVTA 54
Db 217 VPAVVLIGCSTARAHEVLEAARPGQWMLGTGPLPEALPTGRLPGVLALETGQHSIEA 276
Oy 55 VLPRDHPFAFVHGDRHPL-LIAETLEQAMLVFNHAGYGVGVGHPLMTLDYTCCHDHL 113
Db 277 VV-----HDMVELVNAQLSSMAL----- 294

```

```

Oy 114 GVSQGEVLEVEVACSOQK-----FRGGQPVGGQVDMWVRBAGRLAATGTAT 160
Db 295 -VHPRALLPAVAVNCDDDKTGSGSEANGRTLARELGNTSFGQ-----RTGAVWVTGSSQ 346
Oy 161 TRFTSP-QVYRRMGDFATPTASVPGTAP-----VPARAGRTREDDVLSASSQDPTW 213
Db 347 VHVSRHFRVWGLSRDPLGAPAMATVGSWQDQGLDFQGAALR-----VPSPGTQARP 400
Oy 214 RLRYDT--SHPTLFQRPNDH---VPGMLL-----EAAR---QAACLVTPGAP 253
Db 401 KLRVVTLVHPFVFTRESDEGQCPAGQLCLDPTGNTSARLDALFALVNGSV 454

```

## RESULT 15

```

ADN49237
ID ADN49237 standard; protein; 987 AA.
XX
XX AC ADN49237;
XX
XX 01-JUL-2004 (first entry)
XX
XX DE Rat N-methyl-D-aspartate (NMDA) type 3B (NR3B) B4 protein #2.
XX
XX N-methyl-D-aspartate; NMDA type 3B; NR3B; stroke; ischaemia;
XX Creuma; headache; epilepsy; neuropathic pain syndrome;
XX diabetic neuropathy; glaucoma; depression; anxiety; drug addiction;
XX Alzheimer's disease; Huntington's disease; HIV-associated dementia;
XX Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;
XX gene therapy; rat.
XX
XX OS Rattus sp.
XX
XX OS Key Location/Qualifiers
XX FH 1. .17
XX FT Peptide /label= Signal_peptide
XX FT Protein /note= "Rat mature NR3B A2 protein"
XX
XX FT Binding-site /note= "Ligand binding domain"
XX FT /note= "Rat mature NR3B A2 protein"
XX FT Region /note= "Membrane region (M1)"
XX FT /note= "Membrane region (M2)"
XX FT /note= "Membrane region (M3)"
XX FT Binding-site /note= "Ligand binding domain"
XX FT /note= "Membrane region (M4)"
XX
XX PN US2004033500-A1.
XX
XX PD 19-FEB-2004.
XX
XX PF 16-AUG-2002; 2002US-00222772.
XX
XX PR 16-AUG-2002; 2002US-00222772.
XX
XX PA (BURN-) BURNHAM INST.
XX PA (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.
XX
XX PI Lipton SA, Zhang D, Chatterton JE, Sevarino KA, Awobuluyi M;
XX
XX DR WPI; 2004-180043/17.
XX
XX PT N-PSDB; ADN49236.
XX
XX PT New N-methyl-D-aspartate (NMDA) type 3B nucleic acid molecule for
XX preventing or treating conditions associated with abnormal NMDA receptor
XX activation or response to glycine or glutamate, e.g. stroke, anxiety or
XX ischaemia.
XX
XX PS Claim 21; SEQ ID NO 60; 109pp; English.
XX

```



CC The invention relates to N-methyl-D-aspartate (NMDA) type 3B (NR3B)  
CC protein and its corresponding nucleic acid sequence. The invention is  
CC useful for preventing or treating conditions in which inappropriate NMDA  
CC receptor activation or inappropriate responses to glycine or glutamate  
CC are implicated, such as stroke, ischemia, head and spinal trauma,  
CC headache, epilepsy, neuropathic pain syndromes including diabetic  
CC neuropathy, glaucoma, depression and anxiety, drug addiction, Alzheimer's  
CC disease, Huntington's disease, HIV-associated dementia, Parkinson's  
CC disease, multiple sclerosis or amyotrophic lateral sclerosis. It is also  
CC useful in gene therapy. The present sequence is rat NR3B B4 protein.  
XX

XX  
SQ Sequence 987 AA:

Query Match

Best Local Similarity 23.8%; Score 101.5; DB 8; Length 987;

Matches 70; Conservative 31; Mismatches 96; Indels 97; Gaps 13;

QY 1 MEAVVLINSASDANSIEQTALVPVATVHRTVQDAPVSWIPKG-----GDRFSVTA 54  
DB 217 VPAAVLIGCSTARAHVELEAAPPQWMLGTPLPAEALPTTGLPPEVIALGETEQHSLER 276  
QY 55 VLPHDHPFFAPVHGHDPL-LIAETLRQAMLVPHAGYGVPGVHFLMTLDYTCILDHL 113  
DB 277 VV-----HDVVELVAQALSMAL----- 294  
QY 114 GVSGEVAELEVEVACSQK-----FRGGQPVQGOVDNAVRPAGRLAATGTAT 160  
DB 295 -VHPERALPAVVCNCDLKTGSEATGRTIARFLGNTSFOG-----RTGAVWVYTGSSQ 346  
QY 161 TTFSTP-QYRRMRGDFATPTASVPGTAP-----VPAARAGTRDEDEVLSASSQDTW 213  
DB 347 VHVSRHFKVMSLRBDPLGAPAWATVGSWDDGQLDFOPGHAALR-----VSPSGTQARP 400  
QY 214 RLKVDY--SHPTLFORPNDH---VPGMLL-----EAAR---QAACLVTPAP 253  
DB 401 KLRVVTLVVHFPVFTRESDEDDGCPAGQCLDPTNDSARLDALPAALVNGSVP 454

Search completed: July 20, 2005, 22:10:05  
Job time : 127 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 20:11:08 ; Search time 38 Seconds  
(without alignments)  
792.522 Million cell updates/sec

Title: US-10-017-471B-17

Perfect score: 1645  
Sequence: 1 MPEAVVLINSASDANSIEQT.....GHDPGLVLTLSGPAFSG 313

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1.\*  
2: p1r2.\*  
3: p1r3.\*  
4: p1r4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1023.5	62.2	301	2 A32061	af6a protein - Str
2	109.5	5.9	1762	2 T03222	probable polyketid
3	97	5.9	4391	2 A38096	perlecan precursor
4	96.5	5.9	474	1 OMHUIB	alpha-1-B-glycopro
5	94.5	5.7	323	2 H75580	conserved hypotet
6	93	5.7	362	2 T36079	hypothetical prote
7	93	5.7	559	2 T02825	probable membrane
8	93	5.7	786	2 S22155	oncogene 1 (tre-2
9	91.5	5.6	307	2 C75583	hypothetical prote
10	91.5	5.6	407	2 G84309	hypothetical prote
11	91.5	5.6	472	2 F70876	probable pap3 pro
12	91.5	5.6	1634	2 T26517	hypothetical prote
13	91	5.5	912	2 B97566	hypothetical prote
14	91	5.5	1008	2 AH2786	conserved hypotet
15	90.5	5.5	345	2 AH0350	probable DNA-bindi
16	90	5.5	420	2 S30406	hypothetical prote
17	90	5.5	438	2 T34683	hypothetical prote
18	90	5.5	646	2 T35002	probable respirato
19	89.5	5.4	676	2 AG0035	hemin receptor pre
20	89.5	5.4	676	2 T12069	hemin receptor - Y
21	89	5.4	365	2 F96017	hypothetical prote
22	89	5.4	767	2 S41479	DNA-binding protei
23	89	5.4	1460	1 EDBE1F	immediate-early pr
24	89	5.4	2282	2 T42717	DNA-binding protei
25	89	5.4	5149	2 F83345	ATP-dependent RNA
26	88.5	5.4	607	1 S42639	ribozyme polyketi
27	88.5	5.4	5069	2 T17464	CNA peptide synthe
28	88.5	5.4	7463	2 T36248	probable lycopene
29	88	5.3	405	2 T37022	

30	87.5	5.3	172	2 H87253	hypothetical prote
31	87.5	5.3	366	2 C83034	probable oxidoredu
32	87.5	5.3	386	2 T36169	hypothetical prote
33	87.5	5.3	422	2 H75501	DNA/pantothenate m
34	87.5	5.3	626	2 S53871	Pmel 17 protein -
35	87	5.3	559	2 C75286	hypothetical prote
36	87	5.3	606	2 T35710	transferase - Stre
37	87	5.3	1159	2 E70741	probable regulator
38	87	5.3	1241	2 T18311	hypothetical prote
39	86.5	5.3	639	2 F95324	Trag conjugal tran
40	86	5.2	455	2 D84371	phosphomannomutase
41	86	5.2	625	2 S35317	hematopoietic grow
42	86	5.2	4735	2 T17463	ribozyme polyketi
43	85.5	5.2	279	2 AG3540	dihydrodipicolinat
44	85.5	5.2	315	2 S72887	hydroxymethylbilan
45	85.5	5.2	435	2 T03545	probable cobyrimic

#### ALIGNMENTS

##### RESULT 1

A32061  
af6a protein - Streptomyces griseus

C/Species: Streptomyces griseus

C/Date: 31-Aug-1990 #sequence\_revision 31-Aug-1990 #text\_change 09-Jul-2004

C/Accession: A32061

R/Horinouchi, S.; Suzuki, H.; Nishiyama, M.; Beppu, T.

J. Bacteriol. 171, 1206-1210, 1989

A/Title: Nucleotide sequence and transcriptional analysis of the Streptomyces griseus ger

A/Reference number: A32061; MUID:89123125; PMID:2492509

A/Accession: A32061

A/Status: preliminary

A/Molecule type: DNA; mRNA

A/Residues: 1-301 <HOR>

A/Cross-References: UNIROT:P18394; GB:M4250; NID:g153148; PID:AAA26693.1; PID:g153149

C/Superfamily: Streptomyces griseus af6a protein

##### Query Match

Best Local Similarity 66.4%; Pred. No. 3.2e-78;  
Matches 194; Conservative 26; Mismatches 71; Indels 1; Gaps 1;

Qy	23	PVPMAVHRTTRVODAFVSNIPKGRFSYTAVALPHDHPFAPVGHGHDPLLAETLRQ	82
Db	10	PVGIEVHRTTRPEDAPFRNWRILGRDFAVEAVLPHDHPFAPVGDPLVAEAMRQ	69
Qy	83	AAAMVFAAGVGVPGVHFLMT-LDYTHLDHLGVSGEVALVEVAVCSQKFRGGQVVG	141
Db	70	AAAMAFAGVGIPLGVHFLTELDYCHPEHLGVGSEPTIEIGLEVFCSDKMRAGLPAQG	129
Qy	142	QVDAVVRBAGRLAATGATTRFTSPQYRRMRGDFATPTASVPGTAPVPAARAGRTDED	201
Db	130	RVGMVVRBGRLLAATGVAATRFSTPKAYRRMRGDPVEGSLPETAVVPASPRARVED	189
Qy	202	VVLSSASQDTRWLKVDTSPTLFEQRENDHVPGLLLEAARQAACVTGAPAFVPSIGT	261
Db	190	VVLSTGRGEGWELRVDTTRPTLFQRNDHVPGLLLEAARQAACVAGPAGIVPEART	249
Qy	262	RFRVYAFDSPFCWIOATVRGPAPAGLTTVYVTHGDSGLVFTLTSAPASG	313
Db	250	RFRYSEFGSPCWIGAVVOPGADEDTVAVTGHODETFSTVLSGPRAHG	301

##### RESULT 2

T03222  
probable polyketide synthase module 2 - Streptomyces hygroscopicus

C/Species: Streptomyces hygroscopicus

C/Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 09-Jul-2004

C/Accession: T03222

R/Ruan, X.; Staehli, D.; Lax, S.; Katz, L.

Gene 203, 1-9, 1997

A/Title: A second type-I PKS gene cluster isolated from Streptomyces hygroscopicus ATCC

A/Reference number: Z14848; MUID:98085969; PMID:9426000

A:Accession: T03222  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1762 <RNA>  
 A:Cross-references: UNIPROT:Q030480; EMBL:AF007101; NID:G2624946; PIDD:AAAC38062.1; PID:92  
 A:Experimental source: ATCC 29253  
 A:Superfamily: Streptococcus hygroscopicus probable polyketide synthase module 4; 3-oxoac  
 homology; lacyl-carrier-protein S-malonyltransferase homology  
 C:Keywords: carrier protein  
 F:54-454/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>  
 F:550-822/Domain: lacyl-carrier-protein S-malonyltransferase homology <AMT>  
 F:1558-1669/Domain: acyl carrier protein homology <ACP>

Query Match 6.7%; Score 109.5; DB 2; Length 1762;  
 Best Local Similarity 24.6%; Pred. No. 0.92;  
 Matches 93; Conservative 32; Mismatches 162; Indels 91; Gaps 19;

QY 10 SASDANSLEQTLPPPMALVHRTVQDAFPVSWIKGDRPSVAVLHHDHFFAPV--- 66  
 DB 697 SIAMNGPRAVTLSDRRTLLHIAVQLMTKTNWLNVSHAFFSPLMQPILOPFTTTLTL 755  
 QY 67 -HGDRHDPILLIAETLRQAAMLVFHAQGVGVGFHMTLDYCHLDHGVSGEVALELV- 124  
 DB 766 THHPHTPLISMILTPTTHPTTH-----WTQHITAPRYTDTLHLLHHCITTYLEIG 809  
 QY 125 -EVACSOIKFRGCGVQGVQVDAVVR-----AGRLAATG-----TATTFPT 164  
 DB 810 PPTTLTALA-RTTLPTTTHLIPTRRNNEVASTNEALGRVSVGHSDWMLTPTGRT 868  
 QY 165 SPQVRRNRGDPATTAAYVPGAPVPAARAGRTBEDVVLASSQDTRRL---RVDT 220  
 DB 869 SLPTPTFGRRDFWMLHDA-GGGAEEVGAGL-G-TTDPHPLGAVTVAIDGELLISRLSTS 926  
 QY 221 -HPTLFORNDH-----VPGMLLEAARQA-----CLVTGAPAPVP 256  
 DB 927 THFWL---TDHTVNGTVIVPFTALIDLALHAHETDHTVDELVIHPTALHTHPESOS 982  
 QY 257 SIG-----GTRFVRAVEPDS-CPWIOATV-----PPGAPGLTTRVVTGHODGSLVFLTT 305  
 DB 983 TVGAEITDGHRIALHSRTTGTWRHHTGTLSTNGTHPAALSTWSPT---DAHQIDLT 1039  
 QY 306 L-----SGPAFG 313  
 DB 1040 AYQQLADTGLAHYGPAPFG 1057

RESULT 3  
 A38096  
 perlecan precursor - human  
 N:Alternate names: basement membrane heparan sulfate proteoglycan; heparan sulfate prote  
 C:Species: Homo sapiens (man)  
 C:Date: 07-Apr-1994 #sequence revision 07-Apr-1994 #text change 12-Jul-2004  
 A:Accession: A38096; #19256; S77946; A41059; A40306; B33625; A33625; A41736  
 R:Mutidoc: A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V.  
 J. Biol. Chem. 267, 8544-8557, 1992  
 A:Title: Primary structure of the human heparan sulfate proteoglycan from basement mem  
 tor, laminin, neural cell adhesion molecules, and epidermal growth factor.  
 A:Reference number: A38096; MUID:92235084; PMID:1569102  
 A:Accession: A38096  
 A:Molecule type: mRNA  
 A:Residues: 1-4391 <MOR>  
 A:Cross-references: UNIPROT:P98160; GB:M85289; NID:g184426; PIDD:AAA52700.1; PID:g184427  
 R:Kallunki, P.; Trygvason, K.  
 J. Cell Biol. 116, 559-571, 1992  
 A:Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD pro  
 ell adhesion molecules, and epidermal growth factor.  
 A:Reference number: A41736; MUID:92112994; PMID:1730768  
 A:Accession: S19256  
 A:Molecule type: mRNA  
 A:Residues: 157,'59-434,'A',436,'EU',438-449,'Q',451-502,'A',503-792,'K',794-908,'R',  
 71-2979,'H',2981-2994,'G',2996-3167,'T',3169-3240,'R',3242-3426,'R',3428-3631,'Q',3633-3  
 A:Cross-references: EMBL:X62515  
 R:Trygvason, K.

Submitted to the EMBL Data Library, October 1991  
 A:Reference number: S77946  
 A:Accession: S77946  
 A:Molecule type: mRNA  
 A:Residues: 1-57,'D',59-434,'A',436,'EU',438-449,'Q',451-502,'A',503-792,'K',794-908,'R',  
 71-2979,'H',2981-2994,'G',2996-3167,'T',3169-3240,'R',3242-3426,'R',3428-3631,'Q',3633-4  
 A:Cross-references: EMBL:X62515; NID:g29469; PIDD:CAA4373.1; PID:g29470  
 R:Kallunki, P.; Eddy, R.U.; Byers, M.G.; Kestilae, M.; Shows, T.B.; Trygvason, K.  
 Genome 11, 389-396, 1991  
 A:Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the ge  
 A:Reference number: A41059; MUID:92120660; PMID:1685141  
 A:Accession: A41059  
 A:Molecule type: mRNA  
 A:Residues: 'RT',892-908,'R',910-1101,'L',1103-1132,'L',1134-1221,'L',1223-1397 <KA2>  
 A:Cross-references: GB:S76436; NID:g243370; PIDD:AA821121.1; PID:g243371  
 R:Dodge, G.R.; Kovalevsky, I.; Chu, M.L.; Haseell, J.R.; McBride, O.W.; Yi, H.F.; Iozzo, R  
 Genomics 10, 673-680, 1991  
 A:Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cellular  
 A:Reference number: A40306; MUID:91365376; PMID:1679745  
 A:Accession: A40306  
 A:Molecule type: mRNA  
 A:Residues: 1018-1405,'G',1407-1409,'G',1411-1465 <DOD>  
 A:Cross-references: GB:M64283; NID:g184424; PIDD:AAA5269.1; PID:g184425  
 R:Heremans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Caerman, J.J.; van den  
 J. Cell Biol. 109, 3199-3211, 1989  
 A:Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoclonal  
 A:Reference number: A33625; MUID:90078352; PMID:2687294  
 A:Accession: B33625  
 A:Molecule type: protein  
 A:Residues: 1379-1384,'X',1386-1388,'X',1390-1398 <HE2>  
 A:Accession: A33625  
 A:Molecule type: Protein  
 A:Residues: 2166-2171,'X',2173-2175,'X',2177-2185 <HE3>  
 A:Note: peptide potentially matches four different regions of sequence shown  
 C:Genetics:  
 A:Gene: GDB:HSPG2  
 A:Cross-references: GDB:126372; OMIM:142461  
 A:Map position: 1p36.1-1p36.1  
 C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate; transmembra  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:122-4391/Product: perlecan #status predicted <MAT>  
 F:122-193/Domain: I <DOM1>  
 F:194-530/Domain: II <DOM2>  
 F:199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>  
 F:285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
 F:325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>  
 F:368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>  
 F:531-1676/Domain: III <DOM3>  
 F:1159-1206/Domain: laminin-type EGF-like homology <LEG>  
 F:1563-1610/Domain: laminin-type EGF-like homology <EG7>  
 F:1613-1668/Domain: laminin-type EGF-like homology <LEG8>  
 F:1677-2686/Domain: IV <DOM4>  
 F:2007-2034/Domain: transmembrane #status predicted <TM>  
 F:3667-4391/Domain: V <DOM5>  
 F:3845-3880/Domain: EGF homology <EGF1>  
 F:3888-3921/Domain: EGF homology <EGF>  
 F:3993-4106/Domain: laminin G repeat homology <LG2>  
 F:4147-4175/Domain: EGF homology <EGF2>  
 F:4149-4151/Region: motor neuron attachment (L-R-E) motif  
 F:4299-4301/Region: motor neuron attachment (L-R-E) motif  
 F:65,71,76/Binding site: heparan sulfate (Ser) (covalent) #status predicted (asn) (coval  
 F:89,554,1755,2121,3072,3105,3279,3780,3836,4068/Binding site: carbohydrate (Asn) (coval  
 F:2995,3933,4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 5.9%; Score 97; DB 2; Length 4391;  
 Best Local Similarity 22.3%; Pred. No. 31;  
 Matches 101; Conservative 42; Mismatches 133; Indels 176; Gaps 27;

QY 6 VLINSASPN-----SIEGTA-----LPPMALVHRTVQDAFPVSWIKG----- 46  
 DB 2040 VVLSASDASPSPGVKIESSPSVTGGTLLNVCVAGSAHQ-----VTVRRGSLPPT 2095

```

Qy 47 ---GDRSVTVAVLP-----HDPPFPAYHGDHDP 74
Db 2096 QVHSSRLRLPQVSPADSGEYVCRVENSGSPKASITVSVLHGTSGSPYTFVPSTR-PI 2154
Qy 75 LI-----AETL-----RQAMLVPHAGVGVPHFLMTLD- 105
Db 2155 RIBSSSHVAGGQTLDLNCVPGQAHQVTHKRGSGPARHQTHGSLRLHQVTPADSG 2214
Qy 106 -YTCHDLHLGVSG--EVAL------EVENACSQLKFRGQPV----- 139
Db 2215 EYVGHV--VGTSGPLEASVLTIASVIRGPIPRVRISSSSYVAEGTDLSCVVAQA 2272
Qy 140 QGVDMVAVRPAQLAA-----TGATTRF-TSP-----QVTRMRGDPATTPASVPGT--- 186
Db 2273 HAQVTW-YKRGSGSLPARHQVGSRLYIFQASPADAGVYCRASNGMEASITVVTGTGA 2331
Qy 187 -APVPARA-----GTRDEDVVLASAG-QDTWMLR---VDSHPT----- 223
Db 2332 NLAVPAGSTOPRIEPPSSQVAEGQTLDLNCVPGQAHQVTHKRGSLVRHQTHGSL 2391
Qy 224 --LFGQ-----PNDHVPGLLLEAARQAACLV-T-GPAFVPSIGTRFVR-----YAE- 268
Db 2392 LRLVQASPADSGEYVCRVLSVPLEASVLTIERPAGVPALGVTTPVRISSSSQVAEG 2451
Qy 269 --FDSPCWIGATVVRPGPAAGLTVVRVTGHQDG 298
Db 2452 QTLDLNCLV-----AGQAHQVTHKRG 2474

```

## RESULT 4

alpha-1-B-glycoprotein - human

C/Species: Homo sapiens (man)

C/Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 10-Sep-1999

C/Accession: A02113

R:Ichio, N.; Takahashi, N.; Putnam, F.W.

Proc. Natl. Acad. Sci. U.S.A. 83, 2363-2367, 1986

A/Title: Amino acid sequence of human plasma alpha-1B-glycoprotein: homology to the immu

A/Reference number: A02113; MUID:86205955; PMID:3458201

A/Accession: A02113

A/Molecule type: Protein

A/Residues: 1-474 <15H>

C/Comment: The function of this plasma glycoprotein is unknown.

C/Genetics:

A/Gene: GDB:A1B3

A/Cross-references: GDB:119638; OMIM:138670

A/Map position: 19q-19q

C/Superfamily: alpha-1-B-glycoprotein; immunoglobulin homology

C/Keywords: duplication; glycoprotein; plasma

F:21-74/Domain: immunoglobulin homology <IM1>

F:111-163/Domain: immunoglobulin homology <IM2>

F:204-260/Domain: immunoglobulin homology <IM3>

F:297-355/Domain: immunoglobulin homology <IM4>

F:359-451/Domain: immunoglobulin homology <IM5>

F:23,159,342,350/Binding site: carbohydrate (Asn) (covalent) #status experimental

F:28-72,402-449/Disulfide bonds: #status predicted

F:118-161,211-258,304-353/Disulfide bonds: #status experimental

Query Match 5.9%; Score 96.5; DB 1; Length 474;

Best Local Similarity 23.3%; Pred. No. 2.3;

Matches 76; Conservative 37; Mismatches 142; Indels 71; Gaps 14;

```

Qy 8 INASDANSIEQTLAPVPMALVHRTVRQDAPVSWIRPGDRFSTAVLPHDHPFPAVH 67
Db 171 LSEBSATVTIEELAPPPVLMHGB-----SSQVHLHPGNKVTLCVAPLSGVDFQLR 224
Qy 68 GDRHDPPLIAETLQAMLVPHAGVGVPHFLMTLDYTCHDLHLGVSGEVALEVEVA 127
Db 225 GEKE--LLVPRSSSPDRIFFHIN-AVALGCGHNTCYRLHNDNGMSGASAVELLIS 281
Qy 128 CS-----OLKRGQGVQGVDMVAVRPAQLAATGATTRTSQVTRRM 172
Db 282 DETLPAPFSPSPESGRALRLCLAPLEG-ARFALVREDR---GGRVHRFQSPAGTEAL 337

```

```

Qy 173 RGFATPTASVPGTAVPAPARAGTRDEVDVLSASSQDTWRLRVDTSHPTLRQRPNDH 232
Db 338 ---FELNISVADSANSYCVY---DLKPPFGSGASPESELELHVDPPE---RPQLRA 386
Qy 233 PGMLLEAARQAACLVTPGAPFVP-----SIGTRFRFYAFDSCPWQATVR--RGPAGL 287
Db 387 TWGAVLAGDVAVLRCEGPIPDVTPELLRREGTKAVK-----TVRTPGAANL 434
Qy 288 TVVRVTGHQDGLVFLTLGSPAFSG 313
Db 435 -----ELIFV---GPOHAG 445

```

## RESULT 5

H75580

Conserved hypothetical protein - Deinococcus radiodurans (strain R1)

C/Species: Deinococcus radiodurans

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 16-Aug-2004

C/Accession: H75580

R:White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A/Reference number: A75250; MUID:20036896; PMID:10567266

A/Accession: H75580

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-323 <WHI>

A/Cross-references: UNIPROT:Q9RYNT; GB:AE001863; GB:AE001825; NID:96460670; PID:AA1248

A/Experimental source: strain R1

C/Genetics:

A/Gene: DRA0273

A/Map position: 2

C/Superfamily: mult domain homology

Query Match 5.7%; Score 94.5; DB 2; Length 323;

Best Local Similarity 24.3%; Pred. No. 2.1;

Matches 86; Conservative 24; Mismatches 121; Indels 123; Gaps 19;

```

Qy 2 PEAVVLINSASDANSIEQTLAPVPMALVHRTVRQDAPVSWIRPGDRFSTAVLPHDHP 61
Db 35 PRVLVLGSANLARSVNNPFPAPREVEWFRALALDA-----GADLRREV 78
Qy 62 FPAVHDDRRDPLIAETLQAMLVPHAGVGVPHFLMTLDYTCHDLHLGVSGEVAE 121
Db 79 -FRPL-PDRPDADIMADVAVAEIR--GPGASVG-----LVG 113
Qy 122 LEVAVACSQK-FRGGQPVQGVDMVAVRPAQLAATGATTRFTS-----POVYRMR 173
Db 114 FPDASSSYLRWPFPGKRVN-----VPRPGNALDIRALTBGALPAGLPAPVSGAL 167
Qy 174 GDER-TPT-----ASVGTAPVPAPARAGTRDEVDVLSASSQDTWRLR 216
Db 168 GRFAPHTPTCARLREWTAVGGERAALPRTV-----LQERN-LH 206
Qy 217 VDSHPTLRQRPND-----HVGMLLEAARQAACLVTPGAPFVPSIGTRFRVAFED 270
Db 207 VMDRRVLAHRPGIIGGLWELPGRVLPGE-----VLPRGEVLP-GEVPLPGDPTFD 259
Qy 271 SPCWIGAT-----VRPG-PAAGLTVRV-----TGHDGSLVFLTLGSG 308
Db 260 HPAATVTRVVAQIRLGAAPAGGVVPCLEVALSRRLHEDHS-VILARLIG 312

```

## RESULT 6

T36079

hypothetical protein SCR126.09 - Streptomyces coelicolor

C/Species: Streptomyces coelicolor

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

A/Accession: T36079

R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, April 1999

A:Reference number: 221573  
A:Accession: J36079  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-362 <OLI>  
A:Cross-references: UNIPROT:Q9X852; EMBL:AL049630; PIDD:CA040931.1; GSPDB:GN00070; SCOE  
A:Experimental source: strain A3(2)  
C:Genetics:  
A:Gene: SCOE126.09

Query Match 5.7%; Score 93; DB 2; Length 362;  
Best Local Similarity 23.5%; Pred. No. 3.2; Indels 36; Gaps 9;  
Matches 52; Conservative 27; Mismatches 106;

QY 110 LDHLGVSEVAVLEVEVACSLKFRG-----GQVQG-QVDWAVR-----RAGRLAAT 156  
DB 144 LDNVAV--DVAVDAEAVARVLYVGDGABERLGEVAVGAEVLAERWMPMAEGRRAEI 201  
QY 157 GTATRTFTSPQVYRRMRG-----DEATPTASVGTAPVPAPARAGTRD-----ED 201  
DB 202 GLPRDEAASAVVALDAGLAADYAHVSARPPGTLTGFEGETEPVPDGSCTITAH 261  
QY 202 VTLSSAGQDTWRLRVDTSHPTLFQRPNDHVPGLLEAARQAQCLV-TGPAPFVPSIG- 259  
DB 262 VALDCAAAHTAKCTPEYAPNALTRPQREIRALGVGARPPALASTDPAAVRAIAS 321  
QY 260 ---GTRFVRYAEFSDPCWIOATVRPQPAAGLTTVVTGHOD 297  
DB 322 ASQGAELTARGGGLGDFRWLQVGPVDABAL-LVDVADHEE 361

## RESULT 7

202825  
Probable membrane protein L1231.3 [imported] - Leishmania major (strain Friedlin)  
C:Species: Leishmania major  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 09-Jul-2004  
C:Accession: E81455; 102825  
R:Myler, P.; Audleman, L.; devos, T.; Hixson, G.; Kiser, P.; Lemley, C.; Magness, C.;  
Proc. Natl. Acad. Sci. U.S.A. 96, 2902-2906, 1999  
A:Title: Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-c  
A:Reference number: A81455; MUID:991798987; PMID:10077609  
A:Accession: E81455  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-559 <PYL>  
A:Cross-references: UNIPROT:Q00902; GB:AE001274; NID:g3264850; PIDD:AAC24648.1; PID:g226  
A:Experimental source: strain MHOM/IL/81/Friedlin  
C:Genetics:  
A:Gene: L1231.3  
A:Map position: 1  
C:Superfamily: Leishmania major probable membrane protein L1231.3  
C:Keywords: transmembrane protein

Query Match 5.7%; Score 93; DB 2; Length 559;  
Best Local Similarity 21.2%; Pred. No. 5.4; Indels 114; Gaps 14;  
Matches 72; Conservative 39; Mismatches 115;

QY 2 PEAVVLINSASDANSIEOTAL-----PVPMALVHRTVQDAFPVSWIPK 45  
DB 79 PPAVAAASAVTSSAAQSTTVAAAPGVPSGGAASSPVPSPSVLETRQQ-----PA 131  
QY 46 GGDRESVAVLPHDHPFAP-----VHGDRHDPLLIAETLRQAAML-VFHAQYGVV-- 96  
DB 132 DDNRALAFAPARLIALVAPVTLRGRLHDMVSYDAAVHNASILRQLAPYGIHAV 191  
QY 97 -----GTHPLMTLYTCHLD-----HLGVSGEVALEVEVACSLKFRGQGVQ 140  
DB 192 TGAIRGCEPFSLOAEVYLCTLEAATPEVAHVPDAGGLAESD-----R 235  
QY 141 GOVDAAVVRAG-----RLAATGTATRTFTSPQVYR-----RMRGDPAT-----PTAS 182  
DB 236 GAVEPDINDHGDGASGRASAVSTAIAVTDPRTRBARCSRRARAAIQRDALMLRPPS 295

QY 183 VEPAPVPAARAGTRPDEVVLSA-----SSQDTWRLRV----- 217  
DB 296 LPSASSL-AAAASVSRBEDGASVAGGCGHRQOAIRCLGSLRRRYCDRRSAFSSA 354  
QY 218 DTSHTLTFORPNHVPGLLEAARQAQCLVTGPAPFVPS 257  
DB 355 DVSPQVSYRVR-----RALLESIVRCGYIVSAQAQPPPS 388

## RESULT 8

S22155  
oncogene 1 (tre-2 locus) (clone 210) - human  
N:Alternate names: transforming protein  
C:Species: Homo sapiens (man)  
C:Date: 23-Apr-1993 #sequence\_revision 23-Apr-1993 #text\_change 09-Jul-2004  
C:Accession: S57874; S22155  
R:Nakamura, T.; Hillova, J.; Mariage-Samson, R.; Onno, M.; Huebner, K.; Cannizzaro, L.A.;  
Oncogene 7, 733-741, 1992  
A:Title: A novel transcriptional unit of the tre oncogene widely expressed in human cance  
A:Reference number: S57867; MUID:92228503; PMID:1565468  
A:Accession: S57874  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-786 <NAK>  
A:Cross-references: UNIPROT:Q15634; EMBL:X63546; NID:g37329; PIDD:CAA45108.1; PID:g37330

Query Match 5.7%; Score 93; DB 2; Length 786;  
Best Local Similarity 25.5%; Pred. No. 8.3;  
Matches 41; Conservative 21; Mismatches 61; Indels 38; Gaps 8;

QY 145 WAVRAGRLAATGATRTFTSPQVYRRMGDFATPTASVGT---AVPAPARAGT----- 197  
DB 331 WAMNDTVLXHLRASTKLT-----RKQDLPAPARQGSILAPRVVAPASRGKTLCKG 384  
QY 198 -----RDEPVLASASQDPTWRLRVDTSHPTLFQRPNDH-----VPGWLLEAA 241  
DB 385 YRQAPGPPAQFQRFCSAS---PMAKSRSTPCPGAVEDITYPVGTQGVPSLALAQGG 441  
QY 242 RQAA-CLVTGPAPFVPS---IGGTRFVRYAEFSDPCWIOA 277  
DB 442 PQGSMRFLEWMSMRPLPTDLDIGPWFPHY-DFERSCWVA 481

## RESULT 9

C75583  
Hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: C75583  
R:White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; f  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Accession: C75583  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-307 <WHI>  
A:Cross-references: UNIPROT:Q9RYL8; GB:AE001863; GB:AE001825; NID:g6460670; PIDD:AAF1252  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DRA0294  
A:Map position: 2  
C:Superfamily: Deinococcus radiodurans hypothetical protein DRA0294

Query Match 5.6%; Score 91.5; DB 2; Length 307;  
Best Local Similarity 27.2%; Pred. No. 3.5;  
Matches 61; Conservative 12; Mismatches 84; Indels 67; Gaps 11;

QY 113 LGVGEVALEVEVACSLKFRGQPVQGVDM--AVRBAAGRLAA-TGTATRTFTSPQV 169



## RESULT 13

B97566  
hypothetical protein AGR\_C\_3142 [imported] - Agrobacterium tumefaciens (strain C58, Cere  
C/Species: Agrobacterium tumefaciens  
C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 09-Jul-2004  
C/Accession: B97566  
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A/Reference number: A97359; MUID:21608551; PMID:11743194  
A/Accession: B97566  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-912 <KUR>  
A/Cross-references: UNIPROT:O8UEP5; GB:AB007869; PIDD:AAK87483.1; PID:g15156806; GSPDB:C  
C/Genetics:  
A/Gene: AGR\_C\_3142  
A/Map position: circular chromosome

Query Match 5.5%; Score 91; DB 2; Length 912;

Best Local Similarity 20.2%; Pred. No. 15; Mismatches 103; Indels 94; Gaps 12;

Matches 60; Conservative 40; Mismatches 103; Indels 94; Gaps 12;

QY 25 PMALVHRTVQDAFVSWIPKSGDRFSVTAVLPHDHPFAPVHGDRHDLIAETLRQA 84

DB 641 PVDVQRTILPEQLPL-----EGENDADMEAA-----GTPV-GETEDPRLISPEEKAA- 687

QY 85 MLVFNAGVGVGVGHFLMTLDYTCCHDLHGVSG-----EVALLEVAVACSLKFRGG 136

DB 688 -----HNEGAGATGVSPRKRTMTIVKPDGLVAQEDADPA 723

QY 137 QPVGQVDMVAVRAGRLAATGTATRTFTSPQVYRMRGDFA-----TPTASVPGT- 186

DB 724 QPPKAD-----KVALAARPTAKRGEGAVIAASRPVABEQAPQASTPAAKFSTQ 777

QY 187 --APVPAARAGRTDEDVVLSSQODTWRLRVDTSHPTLFQRPNDHVGMLLEAARQA 244

DB 778 SAAPLPARARPS-SQPAVVAVTNOGNV-----RPATTAPA-----QAAQQQ 818

QY 245 ACLVTGAPFPVPSIGTRFVRYAEFDPSCWIOATVRGPAGLTVTVTGHODGSLV 301

DB 819 TAAATPAATSTPSAGG-----YYIQIASLPSQAEAKSYONMSAKFGSVI 863

## RESULT 14

AH2786  
conserved hypothetical protein Atu1710 [imported] - Agrobacterium tumefaciens (strain C5  
C/Species: Agrobacterium tumefaciens  
C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 09-Jul-2004  
C/Accession: AH2786  
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo, I  
erage, G.; Gillet, W.; Grant, C.; Gunttner, D.; Kutyavlin, T.; Levy, R.; Li, M.; McClell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A/Reference number: AB2577; MUID:21608550; PMID:11743193  
A/Accession: AH2786  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-1008 <KUR>  
A/Cross-references: UNIPROT:Q8UEP5; GB:AB008688; PIDD:AAI42710.1; PID:g17740147; GSPDB:C  
C/Genetics:  
A/Experimental source: strain C58 (Dupont)  
A/Gene: Atu1710  
A/Map position: circular chromosome

Query Match 5.5%; Score 91; DB 2; Length 1008;

Best Local Similarity 20.2%; Pred. No. 16;

Matches 60; Conservative 40; Mismatches 103; Indels 94; Gaps 12;

QY 25 PMALVHRTVQDAFVSWIPKSGDRFSVTAVLPHDHPFAPVHGDRHDLIAETLRQA 84

DB 641 PVDVQRTILPEQLPL-----EGENDADMEAA-----GTPV-GETEDPRLISPEEKAA- 687

QY 85 MLVFNAGVGVGVGHFLMTLDYTCCHDLHGVSG-----EVALLEVAVACSLKFRGG 136

DB 784 -----HNEGAGATGVSPRKRTMTIVKPDGLVAQEDADPA 819

QY 137 QPVGQVDMVAVRAGRLAATGTATRTFTSPQVYRMRGDFA-----TPTASVPGT- 186

DB 820 QPPKAD-----KVALAARPTAKRGEGAVIAASRPVABEQAPQASTPAAKFSTQ 873

QY 187 --APVPAARAGRTDEDVVLSSQODTWRLRVDTSHPTLFQRPNDHVGMLLEAARQA 244

DB 874 SAAPLPARARPS-SQPAVVAVTNOGNV-----RPATTAPA-----QAAQQQ 914

QY 245 ACLVTGAPFPVPSIGTRFVRYAEFDPSCWIOATVRGPAGLTVTVTGHODGSLV 301

DB 915 TAAATPAATSTPSAGG-----YYIQIASLPSQAEAKSYONMSAKFGSVI 959

## RESULT 15

AH0350  
Probable DNA-binding protein YP02880 [imported] - Yersinia pestis (strain CO92)

C/Species: Yersinia pestis  
C/Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004  
C/Accession: AH0350  
R/Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titchall, R.W.; Holden, M.T.G.; Prentice, M.B.;  
deno-Tarragge, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; E  
ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett, E  
Nature 413, 523-527, 2001  
A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A/Reference number: AB0001; MUID:21470413; PMID:11586360  
A/Accession: AH0350  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-345 <KUR>  
A/Cross-references: UNIPROT:O8ZCT5; GB:AL590842; PIDD:CA92131.1; PID:g15980847; GSPDB:CN  
C/Genetics:  
A/Gene: YP02880

Query Match 5.5%; Score 90.5; DB 2; Length 345;

Best Local Similarity 24.3%; Pred. No. 4.9;

Matches 59; Conservative 27; Mismatches 90; Indels 67; Gaps 10;

QY 100 FLMTLDYTCCHDLHGVSG-----EVALLEVAVACSLKFRGGQPV----- 139

DB 112 WLMSTWLVLVVLGVTGAMWQHQAQAEIANK-VDQSAQLSQNGQGVPLTDDNSD 170

QY 140 -----OCQVDMVAVRAGRLAATGTATT-----RFTSPQ-----YVRMRGDFA 177

DB 171 AIAPTDAPAPVANOQPVPLTNHSGS-AITNSATTSVVKSTSTPVDYANTNTTTHQDGA 229

QY 178 TPTASVPGTAPVAPARAR-----TRDEDVVLSSAS-----QODTWRLRVDTSHPTLF 225

DB 220 ASAIVSPQVQPPQPTGQPPPLPTADAGVSGSASSVGLVNNFTADCMLOVVDATIGKTLF 289

QY 226 QRPNDHVGMLLEAARQAALVTGPAPFVPSIGTRFVRYAEFDPSCWIOATVRGPAP 285

DB 290 ---SGIQKGAVLVNLA-----GKAPYKLTIGAPALITISYQGNPVLDISKFIKANRYA 338

QY 286 GLT 288

DB 339 RLT 341

Search completed: July 20, 2005, 22:13:04

Job time : 44 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 21:51:49 ; Search time 45 Seconds  
(without alignments)  
519,226 Million cell updates/sec

Title: US-10-017-471B-17

Perfect score: 1645  
Sequence: 1 MPRAVVLINSASDANSIEQT.....GHODGLVLTTLSPAFSG 313

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents AA: \*  
2: /cgn2\_6/ptodata/1/1aa/5A.COMB.pdp: \*  
3: /cgn2\_6/ptodata/1/1aa/5B.COMB.pdp: \*  
4: /cgn2\_6/ptodata/1/1aa/6A.COMB.pdp: \*  
5: /cgn2\_6/ptodata/1/1aa/6B.COMB.pdp: \*  
6: /cgn2\_6/ptodata/1/1aa/6CTUS.COMB.pdp: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	110	6.7	749	US-09-252-991A-20752	Sequence 20752, A
2	104	6.3	660	US-09-252-991A-32871	Sequence 32871, A
3	103	6.3	816	US-09-252-991A-23292	Sequence 23292, A
4	101	6.1	6396	US-09-410-551B-72	Sequence 72, Appl
5	101	6.1	6396	US-09-940-316B-72	Sequence 72, Appl
6	100	6.1	1257	US-09-252-991A-17290	Sequence 17290, A
7	99.5	6.0	632	US-09-252-991A-17148	Sequence 17148, A
8	97.5	5.9	632	US-09-252-991A-17917	Sequence 17917, A
9	97	5.9	4391	US-10-006-011A-2	Sequence 2, Appl
10	94	5.7	239	US-09-252-991A-30746	Sequence 30746, A
11	94	5.7	676	US-09-949-016-7426	Sequence 7426, Ap
12	92.5	5.6	689	US-09-252-991A-31602	Sequence 31602, A
13	92.5	5.6	1490	US-09-252-991A-28442	Sequence 28442, A
14	92	5.6	480	US-09-252-991A-26799	Sequence 26799, A
15	92	5.6	794	US-09-252-991A-28569	Sequence 28569, A
16	91.5	5.5	565	US-09-252-991A-16757	Sequence 16757, A
17	91	5.5	374	US-09-252-991A-23565	Sequence 23565, A
18	90.5	5.5	424	US-09-252-991A-30209	Sequence 30209, A
19	90.5	5.5	548	US-10-096-571-12	Sequence 12, Appl
20	90	5.5	738	US-09-252-991A-30721	Sequence 30721, A
21	89.5	5.4	466	US-09-684-855-113	Sequence 113, App
22	89.5	5.4	466	US-09-684-855-136	Sequence 136, App
23	89.5	5.4	466	US-09-684-855-159	Sequence 159, App
24	89.5	5.4	466	US-09-488-265B-13	Sequence 13, Appl
25	89.5	5.4	487	US-08-868-435-4	Sequence 4, Appl
26	89.5	5.4	487	US-08-744-231-4	Sequence 4, Appl
27	89.5	5.4	487	US-09-636-499-8	Sequence 8, Appl

28	89.5	5.4	487	US-09-273-871A-15	Sequence 15, Appl
29	89.5	5.4	487	US-10-083-452-15	Sequence 15, Appl
30	89.5	5.4	487	US-09-635-504-4	Sequence 4, Appl
31	89.5	5.4	758	US-09-902-540-10176	Sequence 10176, A
32	89	5.4	353	US-09-252-991A-32769	Sequence 32769, A
33	89	5.4	398	US-09-252-991A-27517	Sequence 27517, A
34	89	5.4	2636	US-09-252-991A-25753	Sequence 25753, A
35	88	5.3	1841	US-08-804-227C-6	Sequence 6, Appl
36	88	5.3	4630	US-09-091-609-2	Sequence 2, Appl
37	88	5.3	5215	US-09-105-537-2	Sequence 2, Appl
38	87.5	5.3	316	US-09-902-540-10107	Sequence 10107, A
39	87.5	5.3	577	US-09-252-991A-19178	Sequence 19178, A
40	87	5.3	540	US-09-469-039A-7525	Sequence 7525, Ap
41	87	5.3	589	US-09-252-991A-32979	Sequence 32979, A
42	86.5	5.3	201	US-09-252-991A-18720	Sequence 18720, A
43	86.5	5.3	324	US-09-252-991A-24664	Sequence 24664, A
44	86	5.2	391	US-09-252-991A-20316	Sequence 20316, A
45	86	5.2	426	US-09-902-540-11847	Sequence 11847, A

#### ALIGNMENTS

```
RESULT 1
US-09-252-991A-20752
; Sequence 20752, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: MARC J. RUBENFIELD ET AL.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20752
; LENGTH: 749
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-20752

Query Match      6.7%; Score 110; DB 4; Length 749;
Best Local Similarity 24.7%; Pred. No. 0.011;
Matches 74; Conservative 26; Mismatches 116; Indels 84; Gaps 15;

QY 43 IPKGDRFSTAVLPHDHPFAPVHGDRDPLIA-----ETLRQAMLVFAGYGV 95
DB 465 VPERGDLAGMDVLAHGHEGRLRDAGAGCHDAIVAGCHQLLDLPGHFGANG-ARP 523
QY 96 VGYHFLMTLDYTCGLDHIAGVSEVALEEV-EVACSQLEKFGQGPVQGVQVMAV--RRAGR 152
DB 524 VGLH-----HHGLGGERVFPALPLQVAGQADAKQDHQVODDLAIQRGGX 571
QY 153 LAATGATTT-----RFTSPQVYRRMRGDFATPTASVPG--TAPVPARARTR 198
DB 572 V--EGHATTPPVSPGQWPGSRDAPALREARPGRPSPSPSGRLARAPGP--RHR 626
QY 199 DEDVIVASASQODPWRLRVDTSHPTLRFQRPNHDVPGMLLEAARQACLVTPG----- 251
DB 627 DRCAFPVAGAG-----SSCA--RPRRTPGRLNSAPTRP--ITAPRRRRR 671
QY 252 ----APVPSIGTRFV----RYAEFDSF-----CWIOATVRGPAAGLTTRV 291
DB 672 PARSGVPRPSRGRRRPGRRRYALNDPRARARAGVARRSVCAVATVRTPAGRCGNVR 731

RESULT 2
US-09-252-991A-32871
; Sequence 32871, Application US/09252991A
```

Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 32871  
; LENGTH: 660  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32871

Query Match 6.3%; Score 104; DB 4; Length 660;  
Best Local Similarity 25.9%; Pred. No. 0.041;  
Matches 72; Conservative 17; Mismatches 131; Indels 58; Gaps 12;

QY 34 VQDAPVWIPKGGDRFSTVATLPHDHPFPAVHGDHDDPLIAETLQAAMLVFH-AGY 92  
DB 1 IPSAPVPSWPCWPR-----RPSDDP---GVH-----PLQILPLRRPAAHLAHRPGM 46  
QY 93 GVPVGVHFLMTLDY-TCHLDHLGVSEVALEVEVACSLKFRGGQVPGQVDMNAV-BRA 150  
DB 47 SAARAHRLVLDLGGRAAGRLSAGCGALDLOPPAGREVHVRPFRPQAPGAGRL 106  
QY 151 GRLAATGATATRTFTSPQYRRNRGDFATPTASVPGTAVPAPARAGRTDEVDVLSASSQ 210  
DB 107 QODARTGCLLR---RLLFRGKGPLAQP--AVPPVCGPLPLR----- 144  
QY 211 DTWRARVDTSHFTLFQRPNDRPG-----MLLEAAQAACTVTPAPVPSIGGTFV 264  
DB 145 ---RLRTGSVPRGLDRDPDDIRGAALRLQALRHPRGAFPSAAADGPGSGAGORR-- 199  
QY 265 RYAEFDSPCWIOATVRPG---PAAGLTVTVRVTHODGS 299  
DB 200 -----GRACGIPPRVRPGGGRPAAGADRFRLODQGRGS 232

RESULT 3  
US-09-252-991A-23292  
; Sequence 23292, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23292  
; LENGTH: 816  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23292

Query Match 6.3%; Score 103; DB 4; Length 816;  
Best Local Similarity 31.0%; Pred. No. 0.071;  
Matches 58; Conservative 10; Mismatches 81; Indels 38; Gaps 10;  
QY 135 GGQPVQGVDMVBRAGR-LAATGATATRTFTSPQYR---RWRGDFATPTASVPGTAVP 190  
DB 193 GAAAVHAGADPADPRAGRLSLAAGSRTPLHAQAAHLRFGSRG-----TDPG-APGD 244

QY 191 AARAGRTDEVDVLSASSQDTRWRLRVDTSHPTLFQRPNHVRGMLLEAARQAACLVG 250  
DB 245 PLRQARRRAGQPAAGGGRSPADR-----QRPARRPGRICRA----- 288  
QY 251 PAPVPSIGG---TRFVRVAFDPSPCWIOATVRPGP--AAGLTVTVRVTHODGSVFTTL 306  
DB 289 PGRHHPQLAGGPRR--RTSRVDR---GRLRGAAARRRVRGAGRRIGLCPGGPAP 342  
QY 307 SGPAFSG 313  
DB 343 SGPAVSG 349

RESULT 4  
US-09-410-551B-72  
; Sequence 72, Application US/09410551B  
; Patent No. 6503737  
; GENERAL INFORMATION:  
; APPLICANT: KOSAN BIOSCIENCES, Inc.  
; APPLICANT: REEVES, CHRISTOPHER  
; APPLICANT: CHU, DANIEL  
; APPLICANT: KHOSLA, CHAITAN  
; APPLICANT: SANTO, DANIEL  
; APPLICANT: KU, KAI  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA  
; FILE REFERENCE: 30062-20026.00  
; CURRENT APPLICATION NUMBER: US/09/410,551B  
; CURRENT FILING DATE: 1999-10-01  
; PRIOR APPLICATION NUMBER: US 60/139,650  
; PRIOR FILING DATE: 1999-06-17  
; PRIOR APPLICATION NUMBER: US 60/123,810  
; PRIOR FILING DATE: 1999-03-11  
; PRIOR APPLICATION NUMBER: US 60/102,748  
; PRIOR FILING DATE: 1998-10-02  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 72  
; LENGTH: 6396  
; TYPE: PRT  
; ORGANISM: Streptomyces hygroscopicus  
US-09-410-551B-72

Query Match 6.1%; Score 101; DB 4; Length 6396;  
Best Local Similarity 26.3%; Pred. No. 2.5;  
Matches 97; Conservative 19; Mismatches 127; Indels 126; Gaps 22;

QY 19 QTAIPVPAALVHRRVQ---DAFPVSW-IPKGGDRFSVT-AVLPHDHPFF-----APV 66  
DB 763 QITGPDEVRALHTALADLHVGVADWTLLVIGSGRAPVTLPTYPFOHKDVLRLPSTRADV 822  
QY 67 HGDRHD-----PLLAET-----LRQAAMLVFHA 90  
DB 823 TGAQGEQVAAHLLGNAVALPPTGGVVLTKRLSLASHPWLGHAADVTLVLCGAFLIELAA 882  
QY 91 GYGVPPVGYHFLMTLDYTCCHLDH-----LGVSGEVALEVEVACSLKFRGGQVPG 141  
DB 883 RAGDEVG-----CDLHLELVITETPLVLPATGVVA-VSVHIAEPPDDTGRBAVTVHA 931  
QY 142 QVD-----WAVRAGRLAATGATATRTFTSP-----QYRRMRDGFATPTASV 183  
DB 932 RADSGILM-TRHAGGFLGTAPAPATATDPAPMPAAGPVDVADVYRFE-DIGSYG-- 987  
QY 184 PGTAPVPAA-RAGRTDEVDVLSASSQDTRWRLRVDTSHPTLFQRPNHVRGMLLEAAR 242  
DB 988 PGRGLRAAMRAGDTVYAKVALPDEQSADARF---TLHRA-----LIDPAF 1031  
QY 243 QAACTVGPAPFVPSIGGTRFVRVYAEFDSPCWIOATVRPAGAGLTVTVRVTHODG--SL 300  
DB 1032 QAGALALIDAV-----GGAARLPSPFQD-----VRI-HAAGATRLRLVTVGDGERST 1077  
QY 301 VFLTTLSPG 309

Db 1078 VRMT---GP 1083

## RESULT 5

US-09-940-316B-72  
Sequence 72, Application US/09940316B

Patent No. 6759536  
GENERAL INFORMATION:  
APPLICANT: KOSAN BIOSCIENCES, Inc.  
APPLICANT: REEVES, CHRISTOPHER  
APPLICANT: CHU, DANIEL  
APPLICANT: KHOSLA, CHAITAN  
APPLICANT: SANTI, DANIEL  
APPLICANT: WU, KAI  
TITLE OF INVENTION: POLYKETIDES ENCODING THE FKBP GENE OF THE FK-520 POLYKETIDE SYNTHETASE  
FILE REFERENCE: 30062-20026.11  
CURRENT APPLICATION NUMBER: US/09/940,316B  
PRIOR FILING DATE: 2001-08-27  
PRIOR APPLICATION NUMBER: 09/410,551  
PRIOR FILING DATE: 1999-10-01  
PRIOR APPLICATION NUMBER: US 60/139,650  
PRIOR FILING DATE: 1999-06-17  
PRIOR APPLICATION NUMBER: US 60/123,810  
PRIOR FILING DATE: 1999-03-11  
PRIOR APPLICATION NUMBER: US 60/102,748  
PRIOR FILING DATE: 1998-10-02  
NUMBER OF SEQ ID NOS: 72  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 72  
LENGTH: 6396  
TYPE: PRT  
ORGANISM: Streptomyces hygroscopicus  
US-09-940-316B-72

Query Match 6.1%; Score 101; DB 4; Length 6396;  
Best Local Similarity 26.3%; Pred. No. 2.5;  
Matches 97; Conservative 19; Mismatches 127; Indels 126; Gaps 22;

QY 19 QTALEPMALVHTRVQ---DAFVSW-IRKGRFSVT-AVLPHDHPF-----APV 66  
DB 763 QGTGPRDEVRALHTLAQIHLVRGVAIDMTLVLGGRABVTLPYFQHKDWLRPTSRADV 822  
QY 67 HGRBDH-----PLIAET-----LROAMLVFHA 90  
DB 823 TGAQGEQVAAHPPLGAAVALPBTGSVVLTRLSLASHPMLGHAADVTLPGAAFLDLAA 882  
QY 91 GYGVFVGVHPLMTLDYTCNLDH-----LGVSGEVALEVEVACSQLKFRGQPVQ 141  
DB 883 RAGGEVG-----CDLHLELVETPLVLPATGVAA-VSVEIAEPDDTGRAVYVHA 931  
QY 142 QVD---WAVRAGRLAATGATTTFTSP-----QVYRMGDEPATPTASV 183  
DB 932 RADSSGLM-TRHAGGFGTAPAPATATDPAPWPAEAGPVADVADYDRFE-DIGSYG-- 987  
QY 184 PGTAIPVAA-RAGTRBEDVVLSSSQODTWRLAVDTSHTPLFRPRDHPKMLLEAAR 242  
DB 988 PGFEGFLAAMRAAGTYVAEVALPDEGADAAAF--TLHPP-----LLDAAP 1031  
QY 243 QAACTVGPAPVFSIGTRFVRAEFPSPCMQATVAPGPAAGLTVRVVTHODG--SL 300  
DB 1032 QAGGLAALDAP-----GGAARLPSPFD-----VRI-HAAGATRLRTVGVGDSRST 1077  
QY 301 VFLTTLSP 309  
DB 1078 VRMT---GP 1083

RESULT 6  
US-09-252-991A-17290  
Sequence 17290, Application US/09252991A  
Patent No. 6551795

GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196,136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 17290  
LENGTH: 1257  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17290

Query Match 6.1%; Score 100; DB 4; Length 1257;  
Best Local Similarity 23.3%; Pred. No. 0.28;  
Matches 73; Conservative 26; Mismatches 114; Indels 100; Gaps 13;

QY 3 EAVVLNSASDANSIQTALPVPMALVHRT-RVQDAFPVSWIRKGD--RFSVTAVLP-- 57  
DB 806 DRTAFISAGSVQAQORAFVLGVGLEHAYEBLAEF-----AQGGELLEAVDQVAPRR 860  
QY 58 -----HDHPF---APVHGRDHPPLIAETLRQAMLVFH 89  
DB 861 VALLQAFHQARGETLEGEFAVEHRQRAFRALAYVEGEGHPDLRAFLAVQAVEIHE 920  
QY 90 AGYGVPGVYH-----FLMTLDYTCNLDHLGVSGEVA-----ELEVEVA 127  
DB 921 ARHQVELGQHVDHREARHPQLVQLDNC-ADRLGMGALGVAAQQVQKADGERALDGA 979  
QY 128 CSQKFRGGQPVQ--GVYMAVARAGRLAATGATTTFTSPQYVRRRGRGPAITPVSVP 185  
DB 980 ALALGLEQVEAPRGGVHLGVAVLGIAGVQOHRFV-----G 1019  
QY 186 TAPVPAARAGRTDE-----DVVLSSSQODTWRLRVDTSHPTLFQRPNHVPKMLL 238  
DB 1020 EPPVAVARAADPADRRGAILGGEREVQAGVHQGILARTG-----RADDHVPQQLVQ 1071  
QY 239 EAARQAACLVTPG 251  
DB 1072 VAPLAAAGL--GP 1082

## RESULT 7

US-09-252-991A-17148  
Sequence 17148, Application US/09252991A  
Patent No. 6551795

GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196,136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 17148  
LENGTH: 632  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17148

Query Match 6.0%; Score 99.5; DB 4; Length 632;  
Best Local Similarity 23.7%; Pred. No. 0.11;  
Matches 66; Conservative 26; Mismatches 108; Indels 79; Gaps 12;



QY 183 VPGTAPVPA-----ARAGTRDEVDVLSASSQDTWRLVDTSHPTLFQ 226  
 Db 87 LPRGSPERRRRRLRCRCEQPTDPRRORPGRSDR-----RRHRRRAATRRDRQOPALPP 142  
 QY 227 RPNHVPQMLLEAARQ---AACLVTGPA-----PFVPSIGGT 261  
 Db 143 GRAHRLPGLRAPAATRRRAPAGARRTAPGRQOLPEVPADVNT 184

## RESULT 11

US-09-949-016-7426  
 ; Sequence 7426, Application US/09949016  
 ; Patent No. 681239  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VENTER, J. Craig et al.  
 ; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
 ; FILE REFERENCE: CU001307  
 ; CURRENT APPLICATION NUMBER: US/09/949,016  
 ; PRIOR FILING DATE: 2000-04-14  
 ; PRIOR APPLICATION NUMBER: 60/241,755  
 ; PRIOR FILING DATE: 2000-10-20  
 ; PRIOR APPLICATION NUMBER: 60/237,768  
 ; PRIOR FILING DATE: 2000-10-03  
 ; PRIOR APPLICATION NUMBER: 60/231,498  
 ; PRIOR FILING DATE: 2000-09-08  
 ; NUMBER OF SEQ ID NOS: 207012  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO: 7426  
 ; LENGTH: 676  
 ; TYPE: PRT  
 ; ORGANISM: Human  
 ; US-09-949-016-7426

Query Match 5.7%; Score 94; DB 4; Length 676;  
 Best Local Similarity 28.1%; Pred. No. 0.49;  
 Matches 72; Conservative 16; Mismatches 100; Indels 68; Gaps 17;

QY 7 LINSASDANSIEQ-----TALPVPMALVHRTTRVQDAFPVSWIPKGRSFSTAVL-PPDH 60  
 Db 326 LIHSGSGRRRLPOLRGALPVALV-----PVOGER-----EAAVLHHDH 369  
 QY 61 PFAPVHGDRHDPILIAETLRQAAMLVFHAGVGVPGVYHFLMTLDYT--CHLDH-LGVSG 117  
 Db 370 P-AGPLH-VHR-----VAGAEPLV--AGRGRLGRHALATLPVGRGATLRHQGLGVHG 417  
 QY 118 -----EVALEVEVA--CSQLKFGGQVQGVQVMAVAPAGRLAATGTATTRFT 164  
 Db 418 SAQPVHLPRPHEAQELVTAPRDLHLGARAGDPC-----WEARRLLPLPSAQWDPAPQHP 472  
 QY 165 SP-----QYRRMRGDFATPTASVP-GTAPVPA-ARAGTRDEVDVLSASSQDT 212  
 Db 473 SPUSPSTPISDCACLPAGGRALRTGAPILGALQIPANVLGGVREERQGLGASLPEN 532  
 QY 213 WRLRVDTSHPTLFQRP 228  
 Db 533 LR-SSQSSETLVP RP 547

RESULT 12  
 US-09-252-991A-31602  
 ; Sequence 31602, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 31602  
 ; LENGTH: 689  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-31602

Query Match 5.6%; Score 92.5; DB 4; Length 689;  
 Best Local Similarity 30.4%; Pred. No. 0.73;  
 Matches 45; Conservative 11; Mismatches 59; Indels 33; Gaps 8;

QY 153 LAATGATTTFTSPQVYRRNRGDPATPTASVPGTAPVPAARAGTRDEVDV---VLSASSQ 209  
 Db 48 LARTGSRARSVRPGV---AGHATPRR-PCAGVPAARAGHAGDADVGPRRTGSHQ 102  
 QY 210 QDTWRLAVDTSHPTLFQRPNDVH-PCWMLLEAARQAACLTGPAFPVPSIGGTFRVAYAB 268  
 Db 103 R-----LHPRRRLLPAEAVQPGAGCPHRRPAAPTTGP---PVG-----AA 142  
 QY 269 FDSPCWIO---ATVRPGPAGLTVRV 292  
 Db 143 AGVPAGVRRRPGCAGPFGSGGADATRV 170

## RESULT 13

US-09-252-991A-28442  
 ; Sequence 28442, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; PRIOR FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 28442  
 ; LENGTH: 1490  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 ; US-09-252-991A-28442

Query Match 5.6%; Score 92.5; DB 4; Length 1490;  
 Best Local Similarity 25.2%; Pred. No. 2.3;  
 Matches 69; Conservative 29; Mismatches 99; Indels 77; Gaps 15;

QY 18 EQRALPVPALVHRTVQ---DAFPVSWIPK-----GGDRF-----SVTAVL-- 56  
 Db 290 QQTHAPQGLALHRRARRRVGALFAVQRMQHVGQLGDRIGKEKIDEGRRVAVARIGQ 349  
 QY 57 -----PHDPPFAPVHGDRHDPILIAETLRQA-AMLVFHAGVGVPGVYHFLMT----- 103  
 Db 350 HLAIGGDH-----HGRQVALQLOLDQLRRAEAVEIGHA-----PIQHAYVATLGH 399  
 QY 104 --LDYTCHE-----DHGVSCEVALEVE-VACSQKFRGQV-----VQSQ 142  
 Db 400 RRDPRRLAAADVADGAPABEGADRAFEQLAGHVVRHQQPPQFAPGAGRTQRLVGD 459  
 QY 143 VDMAVRAGRLAATGATTTFTSPQVYRRNRGDPATPTASVPGTAPVPAARAGTRDEVD 202  
 Db 460 VEQVVEPEPATARRALDPQFTAHQVQA---FAVEQEA-GAA--EASRGR-----L 507  
 QY 203 VLSASSQDTWRLVDTSHPTLFQRPNDVHPGWL 236  
 Db 508 GLSETFEDALLFRADADAGVRHQHPOAHQVGV 541

## RESULT 14

US-09-252-991A-26799  
; Sequence 26799, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 26799  
; LENGTH: 480  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26799

Query Match 5.6%; Score 92; DB 4; Length 480;  
Best Local Similarity 28.1%; Pred. No. 0.48;  
Matches 56; Conservative 17; Mismatches 76; Indels 50; Gaps 12;

QY 129 SQLKRGQGVQVQVWVAVRAGRL-----AATGTAATRTFTSPQY-----RMRGDF 176  
DB 204 ARLGRGARPARRRPARFVLRQQLHPAGHPGTGAGHRDQVVGCGAGQPGRGQ 263  
QY 177 ATPTASV-----PGR---APVPAARAGRTDEDVLT-----SASSQDTWRLRYD---T 219  
DB 264 RQPGAPVGRADPGABRSVPVP--PGRHPELRRLROCLRSATSRHSGHGLRADPALR 321  
QY 220 SHPTLFORPNDHVPGLLEAARQA---ACLVGPAPEVPSIGTRFVRYAEFDSPC--W 274  
DB 322 AHPVHFVQPGN--PGAENHSTARHAPDOLRAVRVPLHPHGRKRRPAR-----TPLEW 374  
QY 275 IQATVRP-----GPAAG 286  
DB 375 PRPTARPCLRVAPGGAAPG 393

RESULT 15  
US-09-252-991A-28569  
; Sequence 28569, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 28569  
; LENGTH: 794  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-28569

Query Match 5.6%; Score 92; DB 4; Length 794;  
Best Local Similarity 25.7%; Pred. No. 1;  
Matches 56; Conservative 11; Mismatches 77; Indels 74; Gaps 12;

QY 135 GQPVQGVQVWVAVRAGRLAATGTAATRTFTSPQVYRRMRGDPATPTASVGTAPVP-AAR 193  
DB 102 GSRPICAPA-IASRWAGSARCGSTSTP--TAPATAR-----AASASTPWPAPWPMATN 153  
QY 194 AGRTRD--EDVVISASSQDTWRLRYDTS-----HPTLFORPND---HVPG--MLLEA 240

DB 154 RGRSASFPPTSVSMASRSRTTWBAGTAPARPMYSPTRMATPTSTDRASVPGAVYMTWTS 213  
QY 241 AROA---ACT-----VTGPAPV----- 255  
DB 214 ARSACLSACWRCWPAVPGPAPVPPRPASTPDSTGVGTGKTTSASPIGSRAPSPAS 273  
QY 256 -----PSIGTRFVRYAEFDSPCWQAT--VRPGPAG 286  
DB 274 RCAPRRPGTGATRAISTATSTKPGWSSITPARSGDAPG 311

Search completed: July 20, 2005, 22:13:57  
Job time : 47 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2005, 16:46:04 ; Search time 573.143 Seconds  
(without alignments)  
265.662 Million cell updates/sec

Title: US-10-017-471B-2

Perfect score: 24

Sequence: 1 gtcctgctgcgcgtacacgac 24

Scoring table: IDENTITY NUC  
Gapop 10'-0, Gapext 1.0

Searched: 7173243 seqs, 3172129809 residues

Total number of hits satisfying chosen parameters: 14346486

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database :

Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT NEW PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US10G\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US10G\_PUBCOMB.seq:\*  
20: /cgn2\_6/ptodata/2/pubpna/US10H\_PUBCOMB.seq:\*  
21: /cgn2\_6/ptodata/2/pubpna/US10I\_PUBCOMB.seq:\*  
22: /cgn2\_6/ptodata/2/pubpna/US10\_NEW PUB.seq:\*  
23: /cgn2\_6/ptodata/2/pubpna/US11\_PUBCOMB.seq:\*  
24: /cgn2\_6/ptodata/2/pubpna/US11\_NEW PUB.seq:\*  
25: /cgn2\_6/ptodata/2/pubpna/US60\_NEW PUB.seq:\*  
26: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	22	91.7	24	US-10-017-471A-2	Sequence 2, Appl1
2	19.2	80.0	1224	US-10-437-963-45497	Sequence 45497, A
3	19.2	80.0	4346	US-10-017-471A-19	Sequence 19, Appl1
4	18.8	78.3	1100	US-10-369-493-39801	Sequence 39801, A
5	18.8	78.3	1136	US-10-369-493-39049	Sequence 39049, A
6	18.8	78.3	1136	US-10-369-493-39416	Sequence 39416, A
7	18.8	78.3	2508	US-10-437-963-98196	Sequence 98196, A

C 8	18.8	78.3	4254	17	US-10-282-122A-30494	Sequence 30494, A
C 9	17.8	74.2	365	18	US-10-424-599-23810	Sequence 23810, A
C 10	17.8	74.2	560	20	US-10-425-115-70440	Sequence 70440, A
C 11	17.8	74.2	737	21	US-10-956-157-187	Sequence 187, App
C 12	17.8	74.2	737	21	US-10-956-157-5422	Sequence 5422, Ap
C 13	17.8	74.2	962	13	US-10-027-632-156446	Sequence 156446,
C 14	17.8	74.2	962	13	US-10-027-632-156447	Sequence 156447,
C 15	17.8	74.2	962	13	US-10-027-632-156448	Sequence 156448,
C 16	17.8	74.2	962	17	US-10-027-632-156446	Sequence 156446,
C 17	17.8	74.2	962	17	US-10-027-632-156447	Sequence 156447,
C 18	17.8	74.2	962	17	US-10-027-632-156448	Sequence 156448,
C 19	17.8	74.2	1557	15	US-10-156-761-1939	Sequence 1939, Ap
C 20	17.8	74.2	9025608	15	US-10-156-761-1	Sequence 1, Appl1
C 21	17.6	73.3	433	13	US-10-027-632-127914	Sequence 127914,
C 22	17.6	73.3	4086	19	US-10-437-963-45501	Sequence 45501, A
C 23	17.6	73.3	744	9	US-09-815-242-7762	Sequence 7762, Ap
C 24	17.4	72.5	744	17	US-10-282-122A-30168	Sequence 30168, A
C 25	17.4	72.5	1042	20	US-10-739-930-2160	Sequence 2160, Ap
C 26	17.4	72.5	1348	20	US-10-425-115-60583	Sequence 60583, A
C 27	17.4	72.5	1719	15	US-10-156-761-20779	Sequence 2779, Ap
C 28	17.4	72.5	2553	9	US-09-815-242-7690	Sequence 7690, Ap
C 29	17.4	72.5	2553	17	US-10-282-122A-29994	Sequence 29994, A
C 30	17.4	72.5	9025608	15	US-10-156-761-1	Sequence 1, Appl1
C 31	17.2	71.7	303	19	US-10-437-963-14266	Sequence 14266, A
C 32	17.2	71.7	681	13	US-10-027-632-141887	Sequence 141887,
C 33	17.2	71.7	433	13	US-10-027-632-128283	Sequence 128283,
C 34	17.2	71.7	433	17	US-10-027-632-128283	Sequence 128283,
C 35	17.2	71.7	480	20	US-10-425-115-159020	Sequence 159020,
C 36	17.2	71.7	603	11	US-09-758-759-168	Sequence 168, App
C 37	17.2	71.7	672	19	US-10-767-701-3088	Sequence 3088, Ap
C 38	17.2	71.7	681	13	US-10-027-632-141887	Sequence 141887,
C 39	17.2	71.7	681	13	US-10-027-632-141887	Sequence 141887,
C 40	17.2	71.7	747	15	US-10-156-761-3070	Sequence 3070, Ap
C 41	17.2	71.7	800	13	US-10-027-632-133938	Sequence 133938,
C 42	17.2	71.7	800	17	US-10-027-632-133938	Sequence 133938,
C 43	17.2	71.7	855	18	US-10-425-114-15674	Sequence 15674, A
C 44	17.2	71.7	871	19	US-10-437-963-52721	Sequence 52721, A
C 45	17.2	71.7	954	13	US-10-027-632-101306	Sequence 101306,
C 46	17.2	71.7	954	17	US-10-027-632-101306	Sequence 101306,
C 47	17.2	71.7	963	17	US-10-282-122A-14685	Sequence 14685, A
C 48	17.2	71.7	975	20	US-10-425-115-9817	Sequence 9817, Ap
C 49	17.2	71.7	1417	18	US-10-425-114-2105	Sequence 2105, Ap
C 50	17.2	71.7	1422	17	US-10-282-122A-14757	Sequence 14757, A
C 51	17.2	71.7	1467	15	US-10-156-761-1138	Sequence 1138, Ap
C 52	17.2	71.7	1506	19	US-10-437-963-19687	Sequence 19687, A
C 53	17.2	71.7	1657	17	US-10-374-780A-1560	Sequence 1560, Ap
C 54	17.2	71.7	1699	20	US-10-425-115-24915	Sequence 24915, A
C 55	17.2	71.7	1752	15	US-10-156-761-1235	Sequence 1235, Ap
C 56	17.2	71.7	1821	17	US-10-369-493-47090	Sequence 47090, A
C 57	17.2	71.7	1848	15	US-10-156-761-2720	Sequence 2720, Ap
C 58	17.2	71.7	1866	17	US-10-282-122A-25560	Sequence 25560, A
C 59	17.2	71.7	1920	19	US-10-437-963-56648	Sequence 56648, A
C 60	17.2	71.7	2070	21	US-10-762-107-5	Sequence 5, Appl1
C 61	17.2	71.7	2096	13	US-10-027-632-98759	Sequence 98759, A
C 62	17.2	71.7	2096	17	US-10-027-632-98759	Sequence 98759, A
C 63	17.2	71.7	2189	17	US-10-108-260A-1455	Sequence 1455, Ap
C 64	17.2	71.7	4142	21	US-10-762-107-7	Sequence 7, Appl1
C 65	17.2	71.7	13268	15	US-10-007-926A-47	Sequence 47, Appl1
C 66	17.2	71.7	13268	21	US-10-956-1457-2282	Sequence 2282, Ap
C 67	17.2	71.7	13268	22	US-10-764-425-74	Sequence 74, Appl1
C 68	17.2	71.7	13857	15	US-10-037-270-75	Sequence 75, Appl1
C 69	17.2	71.7	13857	15	US-10-037-270-75	Sequence 75, Appl1
C 70	17.2	71.7	13857	15	US-10-117-722-75	Sequence 75, Appl1
C 71	17.2	71.7	30000	11	US-09-980-217-3	Sequence 3, Appl1
C 72	17.2	71.7	36602	21	US-10-762-107-1	Sequence 1, Appl1
C 73	17.2	71.7	69245	22	US-10-737-082-75	Sequence 75, Appl1
C 74	17.2	71.7	69245	22	US-10-737-082-75	Sequence 75, Appl1
C 75	17.2	71.7	69245	22	US-10-765-790-80	Sequence 80, Appl1
C 76	17.2	71.7	69245	22	US-10-765-790-80	Sequence 80, Appl1
C 77	17.2	71.7	88421	9	US-09-976-059-1	Sequence 1, Appl1
C 78	17.2	71.7	100267	18	US-10-240-425-1470	Sequence 1470, Ap
C 79	17.2	71.7	109519	11	US-09-758-759-1	Sequence 1, Appl1
C 80	17.2	71.7	135638	16	US-10-314-657-1	Sequence 1, Appl1



C 81	17.2	71.7	135638	21	US-10-473-193-1	Sequence 1, Appli	C 154	16.8	70.0	2037	14	US-10-179-510-591	Sequence 591, App
C 82	17	70.8	1656	17	US-10-282-122A-26081	Sequence 26081, A	C 155	16.8	70.0	2037	14	US-10-180-544-591	Sequence 591, App
C 83	16.8	70.0	239	20	US-10-425-115-42958	Sequence 42958, A	C 156	16.8	70.0	2037	14	US-10-180-543-591	Sequence 591, App
C 84	16.8	70.0	277	9	US-09-294-093B-4716	Sequence 4716, Ap	C 157	16.8	70.0	2037	14	US-10-180-546-591	Sequence 591, App
C 85	16.8	70.0	385	19	US-10-767-701-20927	Sequence 20927, A	C 158	16.8	70.0	2037	14	US-10-180-547-591	Sequence 591, App
C 86	16.8	70.0	417	18	US-10-296-115-610	Sequence 610, App	C 159	16.8	70.0	2037	14	US-10-180-549-591	Sequence 591, App
C 87	16.8	70.0	429	17	US-10-291-265-497	Sequence 497, App	C 160	16.8	70.0	2037	14	US-10-180-555-591	Sequence 591, App
C 88	16.8	70.0	524	20	US-10-425-115-53120	Sequence 53120, A	C 161	16.8	70.0	2037	14	US-10-180-559-591	Sequence 591, App
C 89	16.8	70.0	873	17	US-10-369-493-32193	Sequence 32193, A	C 162	16.8	70.0	2037	14	US-10-181-000-591	Sequence 591, App
C 90	16.8	70.0	884	17	US-10-291-265-25	Sequence 25, Appl	C 163	16.8	70.0	2037	14	US-10-183-010-591	Sequence 591, App
C 91	16.8	70.0	1203	9	US-09-815-242-4146	Sequence 4146, Ap	C 164	16.8	70.0	2037	14	US-10-183-012-591	Sequence 591, App
C 92	16.8	70.0	1203	17	US-10-282-122A-7407	Sequence 7407, Ap	C 165	16.8	70.0	2037	14	US-10-184-614-591	Sequence 591, App
C 93	16.8	70.0	1230	9	US-09-785-770A-8	Sequence 8, Appli	C 166	16.8	70.0	2037	14	US-10-184-623-591	Sequence 591, App
C 94	16.8	70.0	1230	17	US-10-409-966-8	Sequence 8, Appli	C 167	16.8	70.0	2037	14	US-10-184-635-591	Sequence 591, App
C 95	16.8	70.0	1254	17	US-10-369-493-40465	Sequence 40465, A	C 168	16.8	70.0	2037	14	US-10-184-637-591	Sequence 591, App
C 96	16.8	70.0	1263	9	US-09-785-770A-7	Sequence 7, Appli	C 169	16.8	70.0	2037	14	US-10-184-646-591	Sequence 591, App
C 97	16.8	70.0	1263	15	US-10-156-761-2153	Sequence 2153, Ap	C 170	16.8	70.0	2037	14	US-10-184-647-591	Sequence 591, App
C 98	16.8	70.0	1263	17	US-10-409-966-7	Sequence 7, Appli	C 171	16.8	70.0	2037	14	US-10-184-652-591	Sequence 591, App
C 99	16.8	70.0	1263	17	US-10-156-761-3467	Sequence 3467, Ap	C 172	16.8	70.0	2037	14	US-10-187-594-591	Sequence 591, App
C 100	16.8	70.0	1278	15	US-10-369-493-24255	Sequence 24255, A	C 173	16.8	70.0	2037	14	US-10-187-596-591	Sequence 591, App
C 101	16.8	70.0	1485	15	US-10-156-761-373	Sequence 373, App	C 174	16.8	70.0	2037	14	US-10-187-745-591	Sequence 591, App
C 102	16.8	70.0	1506	15	US-10-156-761-5829	Sequence 5829, Ap	C 175	16.8	70.0	2037	14	US-10-187-885-591	Sequence 591, App
C 103	16.8	70.0	2037	13	US-10-052-586-591	Sequence 591, App	C 176	16.8	70.0	2037	14	US-10-187-886-591	Sequence 591, App
C 104	16.8	70.0	2037	14	US-10-176-530-591	Sequence 591, App	C 177	16.8	70.0	2037	14	US-10-199-664-591	Sequence 591, App
C 105	16.8	70.0	2037	14	US-10-176-758-591	Sequence 591, App	C 178	16.8	70.0	2037	14	US-10-196-756-591	Sequence 591, App
C 106	16.8	70.0	2037	14	US-10-175-737-591	Sequence 591, App	C 179	16.8	70.0	2037	14	US-10-176-751-591	Sequence 591, App
C 107	16.8	70.0	2037	14	US-10-174-581-591	Sequence 591, App	C 180	16.8	70.0	2037	14	US-10-176-760-591	Sequence 591, App
C 108	16.8	70.0	2037	14	US-10-176-483-591	Sequence 591, App	C 181	16.8	70.0	2037	14	US-10-176-760-591	Sequence 591, App
C 109	16.8	70.0	2037	14	US-10-176-749-591	Sequence 591, App	C 182	16.8	70.0	2037	14	US-10-180-541-591	Sequence 591, App
C 110	16.8	70.0	2037	14	US-10-176-914-591	Sequence 591, App	C 183	16.8	70.0	2037	14	US-10-180-542-591	Sequence 591, App
C 111	16.8	70.0	2037	14	US-10-176-915-591	Sequence 591, App	C 184	16.8	70.0	2037	14	US-10-180-548-591	Sequence 591, App
C 112	16.8	70.0	2037	14	US-10-173-706-591	Sequence 591, App	C 185	16.8	70.0	2037	14	US-10-180-551-591	Sequence 591, App
C 113	16.8	70.0	2037	14	US-10-175-738-591	Sequence 591, App	C 186	16.8	70.0	2037	14	US-10-180-998-591	Sequence 591, App
C 114	16.8	70.0	2037	14	US-10-175-752-591	Sequence 591, App	C 187	16.8	70.0	2037	14	US-10-180-999-591	Sequence 591, App
C 115	16.8	70.0	2037	14	US-10-176-482-591	Sequence 591, App	C 188	16.8	70.0	2037	14	US-10-183-012-591	Sequence 591, App
C 116	16.8	70.0	2037	14	US-10-176-757-591	Sequence 591, App	C 189	16.8	70.0	2037	14	US-10-184-612-591	Sequence 591, App
C 117	16.8	70.0	2037	14	US-10-176-913-591	Sequence 591, App	C 190	16.8	70.0	2037	14	US-10-184-616-591	Sequence 591, App
C 118	16.8	70.0	2037	14	US-10-180-552-591	Sequence 591, App	C 191	16.8	70.0	2037	14	US-10-184-617-591	Sequence 591, App
C 119	16.8	70.0	2037	14	US-10-180-557-591	Sequence 591, App	C 192	16.8	70.0	2037	14	US-10-184-622-591	Sequence 591, App
C 120	16.8	70.0	2037	14	US-10-173-700-591	Sequence 591, App	C 193	16.8	70.0	2037	14	US-10-184-628-591	Sequence 591, App
C 121	16.8	70.0	2037	14	US-10-174-572-591	Sequence 591, App	C 194	16.8	70.0	2037	14	US-10-184-629-591	Sequence 591, App
C 122	16.8	70.0	2037	14	US-10-174-579-591	Sequence 591, App	C 195	16.8	70.0	2037	14	US-10-184-630-591	Sequence 591, App
C 123	16.8	70.0	2037	14	US-10-174-582-591	Sequence 591, App	C 196	16.8	70.0	2037	14	US-10-184-631-591	Sequence 591, App
C 124	16.8	70.0	2037	14	US-10-174-588-591	Sequence 591, App	C 197	16.8	70.0	2037	14	US-10-184-632-591	Sequence 591, App
C 125	16.8	70.0	2037	14	US-10-175-739-591	Sequence 591, App	C 198	16.8	70.0	2037	14	US-10-184-636-591	Sequence 591, App
C 126	16.8	70.0	2037	14	US-10-175-740-591	Sequence 591, App	C 199	16.8	70.0	2037	14	US-10-184-640-591	Sequence 591, App
C 127	16.8	70.0	2037	14	US-10-175-743-591	Sequence 591, App	C 200	16.8	70.0	2037	14	US-10-184-650-591	Sequence 591, App
C 128	16.8	70.0	2037	14	US-10-176-488-591	Sequence 591, App	C 201	16.8	70.0	2037	14	US-10-184-651-591	Sequence 591, App
C 129	16.8	70.0	2037	14	US-10-176-492-591	Sequence 591, App	C 202	16.8	70.0	2037	14	US-10-187-888-591	Sequence 591, App
C 130	16.8	70.0	2037	14	US-10-176-747-591	Sequence 591, App	C 203	16.8	70.0	2037	14	US-10-187-897-591	Sequence 591, App
C 131	16.8	70.0	2037	14	US-10-176-750-591	Sequence 591, App	C 204	16.8	70.0	2037	14	US-10-187-898-591	Sequence 591, App
C 132	16.8	70.0	2037	14	US-10-176-985-591	Sequence 591, App	C 205	16.8	70.0	2037	14	US-10-187-900-591	Sequence 591, App
C 133	16.8	70.0	2037	14	US-10-176-987-591	Sequence 591, App	C 206	16.8	70.0	2037	14	US-10-187-601-591	Sequence 591, App
C 134	16.8	70.0	2037	14	US-10-176-992-591	Sequence 591, App	C 207	16.8	70.0	2037	14	US-10-187-602-591	Sequence 591, App
C 135	16.8	70.0	2037	14	US-10-176-993-591	Sequence 591, App	C 208	16.8	70.0	2037	14	US-10-187-741-591	Sequence 591, App
C 136	16.8	70.0	2037	14	US-10-184-658-591	Sequence 591, App	C 209	16.8	70.0	2037	14	US-10-187-741-591	Sequence 591, App
C 137	16.8	70.0	2037	14	US-10-176-991-591	Sequence 591, App	C 210	16.8	70.0	2037	14	US-10-187-743-591	Sequence 591, App
C 138	16.8	70.0	2037	14	US-10-176-685-591	Sequence 591, App	C 211	16.8	70.0	2037	14	US-10-187-746-591	Sequence 591, App
C 139	16.8	70.0	2037	14	US-10-173-697-591	Sequence 591, App	C 212	16.8	70.0	2037	14	US-10-187-747-591	Sequence 591, App
C 140	16.8	70.0	2037	14	US-10-173-705-591	Sequence 591, App	C 213	16.8	70.0	2037	14	US-10-187-751-591	Sequence 591, App
C 141	16.8	70.0	2037	14	US-10-174-576-591	Sequence 591, App	C 214	16.8	70.0	2037	14	US-10-187-753-591	Sequence 591, App
C 142	16.8	70.0	2037	14	US-10-174-585-591	Sequence 591, App	C 215	16.8	70.0	2037	14	US-10-187-754-591	Sequence 591, App
C 143	16.8	70.0	2037	14	US-10-174-586-591	Sequence 591, App	C 216	16.8	70.0	2037	14	US-10-187-757-591	Sequence 591, App
C 144	16.8	70.0	2037	14	US-10-175-747-591	Sequence 591, App	C 217	16.8	70.0	2037	14	US-10-187-884-591	Sequence 591, App
C 145	16.8	70.0	2037	14	US-10-176-481-591	Sequence 591, App	C 218	16.8	70.0	2037	14	US-10-188-767-591	Sequence 591, App
C 146	16.8	70.0	2037	14	US-10-176-485-591	Sequence 591, App	C 219	16.8	70.0	2037	14	US-10-188-769-591	Sequence 591, App
C 147	16.8	70.0	2037	14	US-10-176-487-591	Sequence 591, App	C 220	16.8	70.0	2037	14	US-10-188-770-591	Sequence 591, App
C 148	16.8	70.0	2037	14	US-10-176-493-591	Sequence 591, App	C 221	16.8	70.0	2037	14	US-10-188-773-591	Sequence 591, App
C 149	16.8	70.0	2037	14	US-10-176-756-591	Sequence 591, App	C 222	16.8	70.0	2037	14	US-10-188-781-591	Sequence 591, App
C 150	16.8	70.0	2037	14	US-10-176-911-591	Sequence 591, App	C 223	16.8	70.0	2037	14	US-10-194-461-591	Sequence 591, App
C 151	16.8	70.0	2037	14	US-10-176-919-591	Sequence 591, App	C 224	16.8	70.0	2037	14	US-10-194-423-591	Sequence 591, App
C 152	16.8	70.0	2037	14	US-10-176-925-591	Sequence 591, App	C 225	16.8	70.0	2037	14	US-10-195-697-591	Sequence 591, App
C 153	16.8	70.0	2037	14	US-10-176-978-591	Sequence 591, App	C 226	16.8	70.0	2037	14	US-10-195-901-591	Sequence 591, App

C 227	16.8	70.0	2037	14	US-10-195-902-591	Sequence 591, App
C 228	16.8	70.0	2037	14	US-10-196-743-591	Sequence 591, App
C 229	16.8	70.0	2037	14	US-10-196-760-591	Sequence 591, App
C 230	16.8	70.0	2037	14	US-10-173-708-591	Sequence 591, App
C 231	16.8	70.0	2037	14	US-10-176-479-591	Sequence 591, App
C 232	16.8	70.0	2037	14	US-10-176-748-591	Sequence 591, App
C 233	16.8	70.0	2037	14	US-10-176-916-591	Sequence 591, App
C 234	16.8	70.0	2037	14	US-10-179-507-591	Sequence 591, App
C 235	16.8	70.0	2037	14	US-10-179-516-591	Sequence 591, App
C 236	16.8	70.0	2037	14	US-10-179-519-591	Sequence 591, App
C 237	16.8	70.0	2037	14	US-10-179-525-591	Sequence 591, App
C 238	16.8	70.0	2037	14	US-10-180-540-591	Sequence 591, App
C 239	16.8	70.0	2037	14	US-10-180-545-591	Sequence 591, App
C 240	16.8	70.0	2037	14	US-10-183-006-591	Sequence 591, App
C 241	16.8	70.0	2037	14	US-10-183-008-591	Sequence 591, App
C 242	16.8	70.0	2037	14	US-10-183-017-591	Sequence 591, App
C 243	16.8	70.0	2037	14	US-10-183-019-591	Sequence 591, App
C 244	16.8	70.0	2037	14	US-10-184-618-591	Sequence 591, App
C 245	16.8	70.0	2037	14	US-10-184-625-591	Sequence 591, App
C 246	16.8	70.0	2037	14	US-10-184-626-591	Sequence 591, App
C 247	16.8	70.0	2037	14	US-10-184-627-591	Sequence 591, App
C 248	16.8	70.0	2037	14	US-10-184-645-591	Sequence 591, App
C 249	16.8	70.0	2037	14	US-10-184-654-591	Sequence 591, App
C 250	16.8	70.0	2037	14	US-10-184-655-591	Sequence 591, App
C 251	16.8	70.0	2037	14	US-10-188-774-591	Sequence 591, App
C 252	16.8	70.0	2037	14	US-10-188-775-591	Sequence 591, App
C 253	16.8	70.0	2037	14	US-10-194-462-591	Sequence 591, App
C 254	16.8	70.0	2037	14	US-10-196-745-591	Sequence 591, App
C 255	16.8	70.0	2037	14	US-10-196-762-591	Sequence 591, App
C 256	16.8	70.0	2037	14	US-10-197-695-591	Sequence 591, App
C 257	16.8	70.0	2037	14	US-10-195-894-591	Sequence 591, App
C 258	16.8	70.0	2037	14	US-10-176-884-591	Sequence 591, App
C 259	16.8	70.0	2037	14	US-10-176-753-591	Sequence 591, App
C 260	16.8	70.0	2037	14	US-10-176-917-591	Sequence 591, App
C 261	16.8	70.0	2037	14	US-10-176-982-591	Sequence 591, App
C 262	16.8	70.0	2037	14	US-10-179-506-591	Sequence 591, App
C 263	16.8	70.0	2037	14	US-10-179-513-591	Sequence 591, App
C 264	16.8	70.0	2037	14	US-10-179-514-591	Sequence 591, App
C 265	16.8	70.0	2037	14	US-10-179-522-591	Sequence 591, App
C 266	16.8	70.0	2037	14	US-10-180-556-591	Sequence 591, App
C 267	16.8	70.0	2037	14	US-10-180-560-591	Sequence 591, App
C 268	16.8	70.0	2037	14	US-10-183-015-591	Sequence 591, App
C 269	16.8	70.0	2037	14	US-10-184-615-591	Sequence 591, App
C 270	16.8	70.0	2037	14	US-10-184-620-591	Sequence 591, App
C 271	16.8	70.0	2037	14	US-10-184-643-591	Sequence 591, App
C 272	16.8	70.0	2037	14	US-10-184-656-591	Sequence 591, App
C 273	16.8	70.0	2037	14	US-10-192-010-591	Sequence 591, App
C 274	16.8	70.0	2037	14	US-10-205-908-591	Sequence 591, App
C 275	16.8	70.0	2037	14	US-10-187-599-591	Sequence 591, App
C 276	16.8	70.0	2037	14	US-10-187-750-591	Sequence 591, App
C 277	16.8	70.0	2037	14	US-10-188-780-591	Sequence 591, App
C 278	16.8	70.0	2037	14	US-10-192-015-591	Sequence 591, App
C 279	16.8	70.0	2037	14	US-10-194-394-591	Sequence 591, App
C 280	16.8	70.0	2037	14	US-10-194-425-591	Sequence 591, App
C 281	16.8	70.0	2037	14	US-10-194-485-591	Sequence 591, App
C 282	16.8	70.0	2037	14	US-10-195-885-591	Sequence 591, App
C 283	16.8	70.0	2037	14	US-10-195-885-591	Sequence 591, App
C 284	16.8	70.0	2037	14	US-10-195-899-591	Sequence 591, App
C 285	16.8	70.0	2037	14	US-10-196-748-591	Sequence 591, App
C 286	16.8	70.0	2037	14	US-10-196-750-591	Sequence 591, App
C 287	16.8	70.0	2037	14	US-10-197-689-591	Sequence 591, App
C 288	16.8	70.0	2037	14	US-10-197-700-591	Sequence 591, App
C 289	16.8	70.0	2037	14	US-10-197-705-591	Sequence 591, App
C 290	16.8	70.0	2037	14	US-10-197-708-591	Sequence 591, App
C 291	16.8	70.0	2037	14	US-10-198-764-591	Sequence 591, App
C 292	16.8	70.0	2037	14	US-10-198-765-591	Sequence 591, App
C 293	16.8	70.0	2037	14	US-10-198-768-591	Sequence 591, App
C 294	16.8	70.0	2037	14	US-10-198-769-591	Sequence 591, App
C 295	16.8	70.0	2037	14	US-10-199-305-591	Sequence 591, App
C 296	16.8	70.0	2037	14	US-10-199-306-591	Sequence 591, App
C 297	16.8	70.0	2037	14	US-10-199-310-591	Sequence 591, App
C 298	16.8	70.0	2037	14	US-10-199-311-591	Sequence 591, App
C 299	16.8	70.0	2037	14	US-10-199-314-591	Sequence 591, App
					US-10-199-317-591	

C 300	16.8	70.0	2037	14	US-10-199-665-591	Sequence 591, App
-------	------	------	------	----	-------------------	-------------------

ALIGNMENTS

RESULT 1

US-10-017--471A-2

Sequence 2, Application US/10017471A

Publication No. US2003012464A1

GENERAL INFORMATION:

APPLICANT: Takano, Eriko

APPLICANT: Bibb, Mervyn

TITLE OF INVENTION: Antibiotic Production

FILE REFERENCE: 0380-P02329US1

CURRENT APPLICATION NUMBER: US/10/017,471A

CURRENT FILING DATE: 2001-10-23

PRIOR APPLICATION NUMBER: US 60/242,561

PRIOR FILING DATE: 2000-10-23

NUMBER OF SEQ ID NOS: 19

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2

LENGTH: 24

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Synthetic

US-10-017-471A-2

Query Match

Best Local Similarity 91.7%; Score 22; DB 15; Length 24;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Qy

1 GTCCGTGTGCCGCGTACGCGSAC 24

|||||

Db

1 GTCCGTGTGCCGCGTACGCGSAC 24

RESULT 2

US-10-437-963-45497

Sequence 45497, Application US/10437963

Publication No. US20040123343A1

GENERAL INFORMATION:

APPLICANT: La Rosa, Thomas J.

APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei

APPLICANT: Wu, Wei

APPLICANT: Boukharov, Andrey A.

APPLICANT: Barbazuk, Brad

APPLICANT: Li, Ping

TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with Plant Improvement

FILE REFERENCE: 38-21(53221)B

CURRENT APPLICATION NUMBER: US/10/437,963

CURRENT FILING DATE: 2003-05-14

NUMBER OF SEQ ID NOS: 204966

SEQ ID NO 45497

LENGTH: 1224

TYPE: DNA

ORGANISM: Oryza sativa

FEATURE:

OTHER INFORMATION: Clone ID: PAT\_MNT4530\_48457C.1

US-10-437-963-45497

Query Match

Best Local Similarity 80.0%; Score 19.2; DB 19; Length 1224;

Matches 78; Conservative 4; Mismatches 2; Indels 0; Gaps 0

Qy

1 GTCCGTGTGCCGCGTACGCGSAC 24

|||||

Db

105 GTCCGTGTGCCGCGTACGCGSAC 128

```

RESULT 3
US-10-017-471A-19
; Sequence 19, Application US/10017471A
; Publication No. US20030124644A1
; GENERAL INFORMATION:
; APPLICANT: Takano, Eriko
; APPLICANT: Bibb, Mervyn
; TITLE OF INVENTION: Antibiotic Production
; FILE REFERENCE: 0380-P02329US1
; CURRENT APPLICATION NUMBER: US/10/017,471A
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: US 60/242,561
; PRIOR FILING DATE: 2000-10-23
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 4346
; TYPE: DNA
; ORGANISM: Streptomyces coelicolor
US-10-017-471A-19

Query Match      80.0%; Score 19.2; DB 15; Length 4346;
Best Local Similarity 75.0%; Pred. No. 42;
Matches 18; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      1 GTCTGTGCGCGGTSAACGCGSAC 24
      |||||:||||:||||:||||:
Db      1250 GTCTGTGATGCCGCGTGACCGCGCAC 1273

RESULT 4
US-10-369-493-39801/c
; Sequence 39801, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 39801
; LENGTH: 1100
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
US-10-369-493-39801

Query Match      78.3%; Score 18.8; DB 17; Length 1100;
Best Local Similarity 70.8%; Pred. No. 83;
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      1 GTCTGTGCGCGGTSAACGCGSAC 24
      |||||:||||:||||:||||:
Db      603 GTCTGTGTTGCCGCGTGACGACGAC 580

RESULT 5
US-10-369-493-39049/c
; Sequence 39049, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.

```

```

; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 39049
; LENGTH: 1136
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
US-10-369-493-39049

Query Match      78.3%; Score 18.8; DB 17; Length 1136;
Best Local Similarity 70.8%; Pred. No. 82;
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      1 GTCTGTGCGCGGTSAACGCGSAC 24
      |||||:||||:||||:||||:
Db      639 GTCTGTGTTGCCGCGTGACGACGAC 616

RESULT 6
US-10-369-493-39416/c
; Sequence 39416, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 39416
; LENGTH: 1136
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
US-10-369-493-39416

Query Match      78.3%; Score 18.8; DB 17; Length 1136;
Best Local Similarity 70.8%; Pred. No. 82;
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      1 GTCTGTGCGCGGTSAACGCGSAC 24
      |||||:||||:||||:||||:
Db      639 GTCTGTGTTGCCGCGTGACGACGAC 616

RESULT 7
US-10-437-963-98196
; Sequence 98196, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

```

FILE REFERENCE: 38-21(5322)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 98196  
LENGTH: 2508  
TYPE: DNA  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_96124C.1  
US-10-437-963-98196

Query Match 78.3%; Score 18.8; DB 19; Length 2508;  
Best Local Similarity 70.8%; Pred. No. 70;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

1 GTCTGTTGGCCGCGTACGCGSAC 24  
456 GTCTGCGAGCGCGCGCGCGCAC 479

## RESULT 8

US-10-282-122A-30494/C  
Sequence 30494, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zykkind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: EUTRA.034A  
CURRENT APPLICATION NUMBER: US/0/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 30494  
LENGTH: 4254  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-10-282-122A-30494

Query Match 78.3%; Score 18.8; DB 17; Length 4254;  
Best Local Similarity 70.8%; Pred. No. 63;

Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

1 GTCTGTTGGCCGCGTACGCGSAC 24  
927 GTCTGTTGGCCGCGCGCGCGCAC 904

## RESULT 9

US-10-424-599-23810/C  
Sequence 23810, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic, David K  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 23810  
LENGTH: 365  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_121501C.1  
US-10-424-599-23810

Query Match 74.2%; Score 17.8; DB 18; Length 365;  
Best Local Similarity 69.6%; Pred. No. 2,8e+02;  
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

1 GTCTGTTGGCCGCGTACGCGSA 23  
260 GTCTGTTGGCCCTTGACGCGGA 238

## RESULT 10

US-10-425-115-70440  
Sequence 70440, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 70440  
LENGTH: 560  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(560)  
OTHER INFORMATION: unsure at all n locations  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_164230C.1  
US-10-425-115-70440

Query Match 74.2%; Score 17.8; DB 20; Length 560;  
Best Local Similarity 69.6%; Pred. No. 2,5e+02;  
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

1 GTCTGTTGGCCGCGTACGCGSA 23  
213 GTCTGTTGGCCGCGTACGCGCA 235

```
RESULT 11
US-10-956-157-187
; Sequence 187, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 187
; LENGTH: 737
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-187

Query Match
Best Local Similarity 74.2%; Score 17.8; DB 21; Length 737;
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTGTGGCCGCGTACGCGGCGC 24
DB 678 TCCTGTGGCCGCGTACGCGGCGC 700

RESULT 12
US-10-956-157-5422/c
; Sequence 5422, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5422
; LENGTH: 737
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-5422

Query Match
Best Local Similarity 74.2%; Score 17.8; DB 21; Length 737;
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 TCCTGTGGCCGCGTACGCGGCGC 24
DB 60 TCCTGTGGCCGCGTACGCGGCGC 38

RESULT 13
US-10-027-632-156446
; Sequence 156446, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; NUMBER OF SEQ ID NOS: 60/218,006
; SOFTWARE: FASTA 3.2
; SEQ ID NO 156446
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-156446

Query Match
Best Local Similarity 76.2%; Score 17.8; DB 13; Length 962;
Matches 16; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
```

```
;; PRIOR APPLICATION NUMBER: US 60/198,676
;; PRIOR FILING DATE: 2000-04-20
;; PRIOR APPLICATION NUMBER: US 60/193,483
;; PRIOR FILING DATE: 2000-03-29
;; PRIOR APPLICATION NUMBER: US 60/185,218
;; PRIOR FILING DATE: 2000-02-24
;; PRIOR APPLICATION NUMBER: US 60/167,363
;; PRIOR FILING DATE: 1999-11-23
;; PRIOR APPLICATION NUMBER: US 60/156,358
;; PRIOR FILING DATE: 1999-09-28
;; PRIOR APPLICATION NUMBER: US 60/146,002
;; PRIOR FILING DATE: 1999-08-09
;; SOFTWARE: FASTA 3.2
;; SEQ ID NO 156446
;; LENGTH: 962
;; TYPE: DNA
;; ORGANISM: Human
US-10-027-632-156446

Query Match
Best Local Similarity 74.2%; Score 17.8; DB 13; Length 962;
Matches 16; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 CTGCTGGCCGCGTACGCGGCGC 24
DB 544 CTGCTGGCCGCGTACGCGGCGC 564

RESULT 14
US-10-027-632-156447
; Sequence 156447, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; NUMBER OF SEQ ID NOS: 60/218,006
; SOFTWARE: FASTA 3.2
; SEQ ID NO 156447
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-156447

Query Match
Best Local Similarity 76.2%; Score 17.8; DB 13; Length 962;
Matches 16; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 CTGCTGGCCGCGTACGCGGCGC 24
DB 544 CTGCTGGCCGCGTACGCGGCGC 564

RESULT 15
US-10-027-632-156448
```

```
Sequence 156448, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 156448
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-156448

Query Match          74.2%; Score 17.8; DB 13; Length 962;
Best Local Similarity 76.2%; Pred. No. 2.3e+02;
Matches 16; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Cy      4 CTGCTGGCCGCGTACGCGCAC 24
      |||:|||||:|||||:|||||:
Db      544 CTGCTGGCCGCGTACGCGCAC 564

RESULT 16
US-10-027-632-156446
; Sequence 156446, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 156446
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-156446
```

```
Query Match          74.2%; Score 17.8; DB 17; Length 962;
Best Local Similarity 76.2%; Pred. No. 2.3e+02;
Matches 16; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Cy      4 CTGCTGGCCGCGTACGCGCAC 24
      |||:|||||:|||||:|||||:
Db      544 CTGCTGGCCGCGTACGCGCAC 564

RESULT 17
US-10-027-632-156447
; Sequence 156447, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 156447
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-156447

Query Match          74.2%; Score 17.8; DB 17; Length 962;
Best Local Similarity 76.2%; Pred. No. 2.3e+02;
Matches 16; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Cy      4 CTGCTGGCCGCGTACGCGCAC 24
      |||:|||||:|||||:|||||:
Db      544 CTGCTGGCCGCGTACGCGCAC 564

RESULT 18
US-10-027-632-156448
; Sequence 156448, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
```

```

; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 156448
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-156448

Query Match
Best Local Similarity 74.2%; Score 17.8; DB 17; Length 962;
Matches 16; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 4 GTGCTGCCGCGTTCACGCCGAC 24
DB 544 GTGCTGCCGCGTTCACGCCGAC 564

RESULT 19
US-10-156-761-1939
; Sequence 1939, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1939
; LENGTH: 1557
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1557)
US-10-156-761-1939

Query Match
Best Local Similarity 74.2%; Score 17.8; DB 15; Length 1557;
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCTGCTGCCGCGTTCACGCCG 23
DB 190 GTCTGCTGCCGCGTTCACGCCG 212

RESULT 20
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
```

```

; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match
Best Local Similarity 74.2%; Score 17.8; DB 15; Length 9025608;
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCTGCTGCCGCGTTCACGCCG 23
DB 2388161 GTCTGCTGCCGCGTTCACGCCG 2388139

RESULT 21
US-10-027-632-127914/c
; Sequence 127914, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 127914
; LENGTH: 413
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-127914

Query Match
Best Local Similarity 73.3%; Score 17.6; DB 13; Length 413;
Matches 17; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTCTGCTGCCGCGTTCACGCCG 24
DB 281 GTCTGCTGCCGCGTTCACGCCG 258

RESULT 22
US-10-027-632-127914/c
; Sequence 127914, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
```



```
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 127914
LENGTH: 413
TYPE: DNA
ORGANISM: Human
US-10-027-632-127914
```

```
Query Match 73.3% Score 17.6; DB 17; Length 413;
Best Local Similarity 70.8%; Pred. No. 3.3e+02;
Matches 17; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```
Oy 1 GTCTGTGCGCCGSGTSCGCSAC 24
Db 281 GTCTGTGCGCCGSGTSCGCSAC 258
```

```
RESULT 23
US-10-437-963-45501
Sequence 45501, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 45501
LENGTH: 4086
TYPE: DNA
ORGANISM: Oryza sativa
FEATURES:
OTHER INFORMATION: Clone ID: PAT_MRT4530_48460C.1
US-10-437-963-45501
```

```
Query Match 73.3% Score 17.6; DB 19; Length 4086;
Best Local Similarity 70.8%; Pred. No. 2.1e+02;
Matches 17; Conservative 4; Mismatches 3; Indels 0; Gaps 0;
```

```
Oy 1 GTCTGTGCGCCGSGTSCGCSAC 24
Db 141 GTCTGTGCGCCGSGTSCGCSAC 164
```

```
RESULT 24
US-09-815-242-7762
Sequence 7762, Application US/09815242
Patent No. US2002006159A1
GENERAL INFORMATION:
APPLICANT: Haeselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyekind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7762
LENGTH: 744
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
FEATURES:
NAME/KEY: CDS
LOCATION: (1)...(744)
US-09-815-242-7762
```

```
Query Match 72.5% Score 17.4; DB 9; Length 744;
Best Local Similarity 71.4%; Pred. No. 3.6e+02;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
```

```
Oy 2 TCCTGTGCGCCGSGTSCGCS 22
Db 602 TCCTGTGCGCCGSGTSCGCS 622
```

```
RESULT 25
US-10-282-122A-30168
Sequence 30168, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haeselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyekind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
```

```

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 30168
; LENGTH: 744
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-10-282-122A-30168

Query Match
Best Local Similarity 72.5%; Score 17.4; DB 17; Length 744;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy      2 TCCTGTGGCGCGTSGACSGS 22
Db      602 TCCTGTGGCGCGTSGCGG 622

RESULT 26
US-10-739-930-2160
; Sequence 2160, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: KOVALIC, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(53377)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO: 2160
; LENGTH: 1042
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: ZEMA-23APR03-CLUSTER281_2
; US-10-739-930-2160

Query Match
Best Local Similarity 72.5%; Score 17.4; DB 20; Length 1042;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy      3 CCTGTGGCGCGTSGACSGSA 23
Db      375 CCTGGCGCGCGTCAACCCGCA 395

RESULT 27
US-10-425-115-60583
; Sequence 60583, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
```

```

; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yinhua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO: 60583
; LENGTH: 1348
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MFT4577_155251C.1
; US-10-425-115-60583

Query Match
Best Local Similarity 72.5%; Score 17.4; DB 20; Length 1348;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy      3 CCTGTGGCGCGTSGACSGSA 23
Db      479 CCTGGCGCGCGTCAACCCGCA 499

RESULT 28
US-10-156-761-2779
; Sequence 2779, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO: 2779
; LENGTH: 1719
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1719)
; US-10-156-761-2779

Query Match
Best Local Similarity 72.5%; Score 17.4; DB 15; Length 1719;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy      4 CTGTGGCGCGTSGACSGSAC 24
Db      1681 CTGTGGCGCGGACGCGGAC 1701

RESULT 29
US-09-815-242-7690/c
; Sequence 7690, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsson, Kari L.
; APPLICANT: Zykkind, Judith W.
```

APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7690  
LENGTH: 2553  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)...(2553)  
US-09-815-242-7690

Query Match 72.5%; Score 17.4; DB 9; Length 2553;  
Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCCTGTGCGCGTACCGC 22  
DB 1550 TCCTGTGCGCGTACCGC 1530

RESULT 30  
US-10-282-122A-29994/C  
Sequence 29994, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zykkind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 29994  
LENGTH: 2553  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-10-282-122A-29994

Query Match 72.5%; Score 17.4; DB 17; Length 2553;  
Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCCTGTGCGCGTACCGC 22  
DB 1550 TCCTGTGCGCGTACCGC 1530

Search completed: July 20, 2005, 22:08:55  
Job time : 605.143 secs

**THIS PAGE IS BLANK**





```
C 247 15.6 65.0 441 4 US-09-270-767-1759 Sequence 3759, Ap
C 248 15.6 65.0 441 4 US-09-270-767-19041 Sequence 19041, A
C 249 15.6 65.0 450 4 US-09-252-991A-6540 Sequence 6540, Ap
C 250 15.6 65.0 450 4 US-09-252-991A-8407 Sequence 8407, Ap
C 251 15.6 65.0 452 4 US-09-513-999C-1145 Sequence 1145, Ap
C 252 15.6 65.0 458 3 US-09-060-756-551 Sequence 551, App
C 253 15.6 65.0 458 4 US-09-670-314-551 Sequence 551, App
C 254 15.6 65.0 462 4 US-09-252-991A-13388 Sequence 13388, A
C 255 15.6 65.0 467 4 US-09-621-976-1973 Sequence 1973, Ap
C 256 15.6 65.0 467 4 US-09-621-976-17391 Sequence 17391, A
C 257 15.6 65.0 471 4 US-09-513-999C-36158 Sequence 36158, A
C 258 15.6 65.0 516 4 US-09-252-991A-508 Sequence 508, App
C 259 15.6 65.0 537 4 US-09-252-991A-5627 Sequence 5627, App
C 260 15.6 65.0 547 4 US-09-513-999C-3905 Sequence 3905, Ap
C 261 15.6 65.0 573 4 US-09-902-540-4384 Sequence 4384, Ap
C 262 15.6 65.0 600 4 US-09-894-844-41 Sequence 41, Appl
C 263 15.6 65.0 601 4 US-09-949-016-29248 Sequence 29248, A
C 264 15.6 65.0 601 4 US-09-949-016-45231 Sequence 45231, A
C 265 15.6 65.0 601 4 US-09-949-016-58791 Sequence 58791, A
C 266 15.6 65.0 601 4 US-09-949-016-58792 Sequence 58792, A
C 267 15.6 65.0 601 4 US-09-949-016-62891 Sequence 62891, A
C 268 15.6 65.0 601 4 US-09-949-016-202923 Sequence 202923,
C 269 15.6 65.0 624 4 US-09-252-991A-6551 Sequence 6551, Ap
C 270 15.6 65.0 648 4 US-09-107-433-1573 Sequence 1573, Ap
C 271 15.6 65.0 672 4 US-09-252-991A-9502 Sequence 9502, Ap
C 272 15.6 65.0 672 4 US-09-949-016-465 Sequence 465, App
C 273 15.6 65.0 677 4 US-09-949-016-4472 Sequence 4472, Ap
C 274 15.6 65.0 678 4 US-09-902-540-9065 Sequence 9065, Ap
C 275 15.6 65.0 723 4 US-09-252-991A-13839 Sequence 13839, A
C 276 15.6 65.0 723 4 US-09-902-540-7404 Sequence 7404, Ap
C 277 15.6 65.0 744 4 US-09-489-039A-6647 Sequence 6647, Ap
C 278 15.6 65.0 765 4 US-09-252-991A-6511 Sequence 6511, Ap
C 279 15.6 65.0 788 3 US-09-142-623-10 Sequence 10, Appl
C 280 15.6 65.0 795 4 US-09-902-540-6542 Sequence 6542, Ap
C 281 15.6 65.0 822 4 US-09-252-991A-8940 Sequence 8940, Ap
C 282 15.6 65.0 837 4 US-09-902-540-3467 Sequence 3467, Ap
C 283 15.6 65.0 843 4 US-09-902-540-8507 Sequence 8507, Ap
C 284 15.6 65.0 845 4 US-09-902-540-8861 Sequence 8861, Ap
C 285 15.6 65.0 864 4 US-09-252-991A-6520 Sequence 6520, Ap
C 286 15.6 65.0 867 4 US-09-252-991A-13023 Sequence 13023, A
C 287 15.6 65.0 875 4 US-09-902-540-3349 Sequence 3349, Ap
C 288 15.6 65.0 894 4 US-09-252-991A-11968 Sequence 11968, A
C 289 15.6 65.0 897 4 US-09-949-016-2619 Sequence 2619, Ap
C 290 15.6 65.0 897 4 US-09-949-016-2900 Sequence 2900, Ap
C 291 15.6 65.0 908 3 US-08-660-645A-4 Sequence 4, Appl
C 292 15.6 65.0 908 3 US-09-298-718-4 Sequence 4, Appl
C 293 15.6 65.0 908 3 US-09-546-969-4 Sequence 4, Appl
C 294 15.6 65.0 908 4 US-09-547-267-4 Sequence 4, Appl
C 295 15.6 65.0 918 4 US-09-252-991A-8355 Sequence 8355, Ap
C 296 15.6 65.0 921 4 US-09-252-991A-12978 Sequence 12978, A
C 297 15.6 65.0 954 4 US-09-489-039A-2597 Sequence 2597, Ap
C 298 15.6 65.0 975 4 US-09-252-991A-12499 Sequence 12499, A
C 299 15.6 65.0 984 3 US-08-459-046-1 Sequence 1, Appl
C 300 15.6 65.0 984 3 US-08-513-974B-57 Sequence 57, Appl
```

## ALIGNMENTS

```
RESULT 1
US-09-252-991A-10060
; Sequence 10060, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
```

```
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10060
; LENGTH: 669
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10060
Query Match 78.3%; Score 18.8; DB 4; Length 669;
Best Local Similarity 70.8%; Pred. No. 1.2e+02;
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
```

```
Cy 1 GTCCTGTCGCGCGTACGCGSAC 24
Db 252 GTCCTGTCGCGCGCACCGCGCAC 275
```

```
RESULT 2
US-09-252-991A-9773/c
; Sequence 9773, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9773
; LENGTH: 810
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9773
Query Match 78.3%; Score 18.8; DB 4; Length 810;
Best Local Similarity 70.8%; Pred. No. 1.2e+02;
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
```

```
Cy 1 GTCCTGTCGCGCGTACGCGSAC 24
Db 634 GTCCTGTCGCGCGCACCGCGCAC 611
```

```
RESULT 3
US-09-252-991A-8162/c
; Sequence 8162, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 8162
; LENGTH: 1338
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-8162
Query Match 78.3%; Score 18.8; DB 4; Length 1338;
Best Local Similarity 70.8%; Pred. No. 1.2e+02;
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
```



```

QY      1  GTCCTGTTGGCCSGTSAACSGSAC  24
          ||||:||||:||||:||||
Db      999  GTCCGGCTGGCCGGTCAACCCAC  976

```

RESULT 4  
US-09-902-540-4394/c  
; Sequence 4394, Application US/09902540  
; Patent No. 6827447

```

; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Hinkler, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 4394
; LENGTH: 2403
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-4394

```

Query Match	78.3%	Score 18.8	DB 4	Length 2403
Best Local Similarity	70.8%	Pred. No. 1.2e+02		
Matches 17; Conservative	5	Mismatches 2	Indels 0	Gaps 0

QY	1	GTCTGTGGCCSGTSACSGSAC	24
		:    :    :    :	
Db	777	GTCTGCGAGCCCGTACCCGCAC	754

RESULT 5  
US-09-252-991A-9867/c  
; Sequence 9867, Application US/09252991A

```

: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,150
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 9867
: LENGTH: 4248
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
: US-09-252-991A-9867

```

Query Match	78.3%;	Score 18.8;	DB 4;	Length 4248;
Best Local Similarity	70.8%;	Pred. No.1.2e+02;		
Matches 17;	Conservative 5;	Mismatches 2;	Indels 0;	Gaps 0

OY	1	GTCTG	STGGCC	SGT	SAC	SG	SAC	24
			:	:	:	:	:	
Db	975	GTCTG	CTGGCC	GGC	AC	CCG	CAC	952

RESULT 6  
US-08-804-227C-13  
; Sequence 13, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

ADDRESS: THOMAS G. FLANN 150  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZTD: 46395

```
?  
?  
?   MODEL: 70203  
?  
?   COMPUTER READABLE FORM:  
?  
?     MEDIUM TYPE: Floppy disk  
?  
?   COMPUTER: IBM Compatible  
?  
?   OPERATING SYSTEM: MS-DOS  
?  
?   SOFTWARE: ASCII(DOS) Text only  
?  
?   CURRENT APPLICATION DATA: /usr/local
```

APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784

```
; REFERENCE/DOCKET NUMBER: A  
; TELECOMMUNICATION INFORMATION:  
REV EDIOWT 317-376-3450
```

TELEPHONE: 317-276-2439  
; INFORMATION FOR SEQ ID NO: 13  
; SEQUENCE CHARACTERISTICS:

SEQUENCE CHARACTERISTICS:  
LENGTH: 13987 base pairs  
TYPE: nucleic acid

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

```

;
;      NAME/KEY: CDS
;      LOCATION: 350 13987

```

US-08-804-227C-13

Query Match	78.3%	Score 18.8	DB 2	Length 13987
Best Local Similarity	70.8%	Pred. No. 1.2e+02		
Matches 17, Conservative	5	Mismatches 2	Indels 0	Gaps 0

QY	1	GTCTGTGTGGCCSGTSACSGSAC	24
		:           :	
Db	8324	GACCTGTGTGGCCGCTGACGGGCAC	8347

RESULT 7  
US-09-902-540-1217  
: Sequence 1217, Application US/09902540

GENERAL INFORMATION: Barry S.  
APPLICANT: Goldman, Gregory J.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Hinkle, Gregory J.

1  
 2 APPLICANT: Slater, Steven C.  
 3 TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
 4 FILE REFERENCE: 38-10 (15849) B  
 5 CURRENT APPLICATION NUMBER: US/09/902,540  
 6 CURRENT FILING DATE: 2001-07-10  
 7 PRIOR APPLICATION NUMBER: 60/217, 883  
 8 PRIOR FILING DATE: 2000-07-10  
 9 NUMBER OF SEQ ID NOS: 16825  
 10

```

? SEQ ID NO 1217
? LENGTH: 292272
? TYPE: DNA
? ORGANISM: Myxococcus xanthus
US-09-902-540-1217

```

Query Match	Score	DB 4	Length
Best Local Similarity	78.3%;	70.8%;	29272;

Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCCTGTTGCCGCTGACGCGCAC 24  
|||||:|||||:|||||:|||||:  
Db 20173 GTCCTGCCAGCCGCTGACGCGCAC 20196

RESULT 8  
US-08-804-227C-7  
Sequence 7, Application US/08804227C  
Patent No. 5876991  
GENERAL INFORMATION:  
APPLICANT: Dehoff, Bradley S.  
APPLICANT: Kuhseos, Stuart A.  
APPLICANT: Rostock, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCII(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44377 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 350..14002  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 14046..20036  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 20110..31284  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31329..36071  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 36155..41830  
US-08-804-227C-7

Query Match 78.3%; Score 18.8; DB 2; Length 44377;  
Best local Similarity 70.8%; Pred. No. 1.2e+02;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCCTGTTGCCGCTGACGCGCAC 24  
|||||:|||||:|||||:|||||:  
Db 8339 GACCTGTGCGCGGTGACGCGCAC 8362

RESULT 9  
US-08-804-198-1  
Sequence 1, Application US/08804198  
Patent No. 5945320  
GENERAL INFORMATION:  
APPLICANT: Burgett, Stanley G.  
APPLICANT: Kuhseos, Stuart A.  
APPLICANT: Rao, Nagarsja R.  
APPLICANT: Richardson, Mark A.  
APPLICANT: Rostock, Paul R., Jr.  
TITLE OF INVENTION: PLATENOIDE SYNTHASE GENE  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PAUL R. CANTRELL 1138  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: Macintosh 7.0  
SOFTWARE: Microsoft Word 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,198  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: CANTRELL, PAUL R.  
REGISTRATION NUMBER: 36,470  
REFERENCE/DOCKET NUMBER: P9113  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-3885  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 44377 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 350..14002  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 14046..20036  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 20110..31284  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 31329..36071  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 36155..41830  
US-08-804-198-1

Query Match 78.3%; Score 18.8; DB 2; Length 44377;  
Best local Similarity 70.8%; Pred. No. 1.2e+02;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCCTGTTGCCGCTGACGCGCAC 24  
|||||:|||||:|||||:|||||:  
Db 8339 GACCTGTGCGCGGTGACGCGCAC 8362

RESULT 10  
US-09-902-540-3008  
Sequence 3008, Application US/09902540  
Patent No. 6833447  
GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.

```

; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; NUMBER OF SEQ ID NOS: 16825
; LENGTH: 432
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-3008

Query Match          75.0%; Score 18; DB 4; Length 432;
Best Local Similarity 75.0%; Pred. No. 2.6e+02;
Matches 15; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCTGTGGCGCGTSAACGCGS 22
    |||||:|||||:|||||:|||||:
DB 75 CCTGTGGCGCGTSAACGCGC 94

RESULT 11
US-09-902-540-636/c
; Sequence 636, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 636
; LENGTH: 3092
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-636

Query Match          75.0%; Score 18; DB 4; Length 3092;
Best Local Similarity 75.0%; Pred. No. 2.5e+02;
Matches 15; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCTGTGGCGCGTSAACGCGS 22
    |||||:|||||:|||||:|||||:
DB 3017 CCTGTGGCGCGTSAACGCGC 2998

RESULT 12
US-09-252-991A-1308/c
; Sequence 1308, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1308
```

```

; LENGTH: 375
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1308

Query Match          74.2%; Score 17.8; DB 4; Length 375;
Best Local Similarity 69.6%; Pred. No. 3.1e+02;
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCTGTGGCGCGTSAACGCGSA 23
    |||||:|||||:|||||:|||||:
DB 117 GTCTGTGGCGCGTSAACGCGGA 95

RESULT 13
US-09-252-991A-1157
; Sequence 1157, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1157
; LENGTH: 681
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1157

Query Match          74.2%; Score 17.8; DB 4; Length 681;
Best Local Similarity 69.6%; Pred. No. 3.1e+02;
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCTGTGGCGCGTSAACGCGSA 23
    |||||:|||||:|||||:|||||:
DB 610 GTCTGTGGCGCGTSAACGCGGA 632

RESULT 14
US-09-252-991A-5982
; Sequence 5982, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5982
; LENGTH: 492
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5982

Query Match          72.5%; Score 17.4; DB 4; Length 492;
Best Local Similarity 71.4%; Pred. No. 4.5e+02;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCTGTGGCGCGTSAACGCGS 22
    |||||:|||||:|||||:|||||:
```

Db 232 TCCTGCTGCGCGATCACCCGC 252

RESULT 15

US-09-252-991A-339/c  
; Sequence 339, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 339  
; LENGTH: 654  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-339

Query Match 72.5%; Score 17.4; DB 4; Length 654;  
Best Local Similarity 71.4%; Pred. No. 4.5e+02;

Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 548 TCCTGCTGCGCGCGGCGCG 528

Query 2 TCCTGCTGCGCGGCGCGCG 22

Db 548 TCCTGCTGCGCGCGGCGCG 528

RESULT 16

US-09-252-991A-330  
; Sequence 330, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 330  
; LENGTH: 759  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-330

Query Match 72.5%; Score 17.4; DB 4; Length 759;  
Best Local Similarity 71.4%; Pred. No. 4.5e+02;

Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 617 TCCTGCTGCGCGGCGCGCG 637

Query 2 TCCTGCTGCGCGGCGCGCG 22

Db 617 TCCTGCTGCGCGGCGCGCG 637

RESULT 17

US-09-489-039A-3425  
; Sequence 3425, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 3425  
; LENGTH: 843  
; TYPE: DNA  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-3425

Query Match 72.5%; Score 17.4; DB 4; Length 843;  
Best Local Similarity 71.4%; Pred. No. 4.5e+02;

Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 337 CTGCTGCGCGATGACCGGAC 357

Query 4 CTGCTGCGCGGCGGCGGAC 24

Db 337 CTGCTGCGCGATGACCGGAC 357

RESULT 18

US-09-252-991A-319  
; Sequence 319, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 319  
; LENGTH: 1026  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: unseq  
; LOCATION: (962)  
; OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.  
US-09-252-991A-319

Query Match 72.5%; Score 17.4; DB 4; Length 1026;  
Best Local Similarity 71.4%; Pred. No. 4.5e+02;

Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Db 78 TCCTGCTGCGCGGCGCGCG 98

Query 2 TCCTGCTGCGCGGCGCGCG 22

Db 78 TCCTGCTGCGCGGCGCGCG 98

RESULT 19

5212296-17  
; Patent No. 5212296  
; APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LIETO, KENNETH  
; J.; O'KEEFE, DANIEL P.; JOWER, CHARLES A.; ROMESSER, JAMES A.  
; TEPERMAN, JAMES M.  
; TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING  
; CYTOCHROMES  
; NUMBER OF SEQUENCES: 19  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/569,781  
; FILING DATE: 23-AUG-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 464,499  
; FILING DATE: 12-JAN-1990  
; APPLICATION NUMBER: 405,605  
; FILING DATE: 11-SEP-1989

```
; SEQ ID NO:17:
; LENGTH: 1212

Query Match
Best Local Similarity 71.4%; Pred. No. 4.5e+02;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCCTGCTGGCGCGTGTACCGCGC 22
DB 148 TCCTGCTGGCGCGTGTACCGCGC 168

RESULT 20
5212296-17
; Patent No. 5212296
; APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH
; J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.
; PEPERMAN, JAMES M.
; TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING
; CYTOCHROMES
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/569,781
; FILING DATE: 23-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 464,499
; FILING DATE: 12-JAN-1990
; APPLICATION NUMBER: 405,605
; FILING DATE: 11-SEP-1989
; SEQ ID NO:17:
; LENGTH: 1212

Query Match
Best Local Similarity 71.4%; Pred. No. 4.5e+02;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCCTGCTGGCGCGTGTACCGCGC 22
DB 148 TCCTGCTGGCGCGTGTACCGCGC 168

RESULT 21
US-09-252-991A-5965
; Sequence 5965, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5965
; LENGTH: 1578
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5965

Query Match
Best Local Similarity 72.5%; Score 17.4; DB 4; Length 1578;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCCTGCTGGCGCGTGTACCGCGC 22
DB 434 TCCTGCTGGCGCGTGTACCGCGC 454

; SEQ ID NO:17:
; LENGTH: 1212

Query Match
Best Local Similarity 71.4%; Pred. No. 4.5e+02;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCCTGCTGGCGCGTGTACCGCGC 22
DB 148 TCCTGCTGGCGCGTGTACCGCGC 168

RESULT 22
US-09-252-991A-11053
; Sequence 11053, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11053
; LENGTH: 1827
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11053

Query Match
Best Local Similarity 72.5%; Score 17.4; DB 4; Length 1827;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCTGCTGGCGCGCGTGTACCGCGA 23
DB 228 CCTGCTGGCGCGCGTGTACCGCGA 248

RESULT 23
US-09-252-991A-11201/c
; Sequence 11201, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11201
; LENGTH: 1929
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11201

Query Match
Best Local Similarity 72.5%; Score 17.4; DB 4; Length 1929;
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 3 CCTGCTGGCGCGCGTGTACCGCGA 23
DB 1660 CCTGCTGGCGCGCGTGTACCGCGA 1640

RESULT 24
US-09-252-991A-10982
; Sequence 10982, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
```

CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 10982  
LENGTH: 1962  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-10982

Query Match 72.5%; Score 17.4; DB 4; Length 1962;  
Best Local Similarity 71.4%; Pred. No. 4.4e+02;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 3 TCCTGTGGCCGCGTACCGCGA 23  
Db 289 CCTGCTGGCGCGACCGCGA 309

RESULT 25  
5212296-8  
PATENT NO. 5212296  
APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH  
J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.  
TEPPERMAN, JAMES M.  
TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING  
CYTOCHROMES  
NUMBER OF SEQUENCES: 19  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/569,781  
FILING DATE: 23-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 464,499  
FILING DATE: 12-JAN-1990  
APPLICATION NUMBER: 405,605  
FILING DATE: 11-SEP-1989  
SEQ ID NO: 8  
LENGTH: 1998  
5212296-8

Query Match 72.5%; Score 17.4; DB 6; Length 1998;  
Best Local Similarity 71.4%; Pred. No. 4.4e+02;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCCTGTGGCCGCGTACCGCG 22  
Db 342 TCCTGTGGCGGTGACCGCG 362

RESULT 26  
5212296-8  
PATENT NO. 5212296  
APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH  
J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.  
TEPPERMAN, JAMES M.  
TITLE OF INVENTION: EXPRESSION OF HERBICIDE METABOLIZING  
CYTOCHROMES  
NUMBER OF SEQUENCES: 19  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/569,781  
FILING DATE: 23-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 464,499  
FILING DATE: 12-JAN-1990  
APPLICATION NUMBER: 405,605  
FILING DATE: 11-SEP-1989  
SEQ ID NO: 8  
LENGTH: 1998  
5212296-8

Query Match 72.5%; Score 17.4; DB 6; Length 1998;

Best Local Similarity 71.4%; Pred. No. 4.4e+02;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
Qy 2 TCCTGTGGCCGCGTACCGCG 22  
Db 342 TCCTGTGGCGGTGACCGCG 362

RESULT 27  
US-09-252-991A-5908/C  
Sequence 5908, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 5908  
LENGTH: 2469  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-5908

Query Match 72.5%; Score 17.4; DB 4; Length 2469;  
Best Local Similarity 71.4%; Pred. No. 4.4e+02;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCCTGTGGCCGCGTACCGCG 22  
Db 786 TCCTGTGGCGGTACCGCG 766

RESULT 28  
US-09-252-991A-5922/C  
Sequence 5922, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 5922  
LENGTH: 3135  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-5922

Query Match 72.5%; Score 17.4; DB 4; Length 3135;  
Best Local Similarity 71.4%; Pred. No. 4.4e+02;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCCTGTGGCCGCGTACCGCG 22  
Db 2132 TCCTGTGGCGGTACCGCG 2112

RESULT 29  
US-09-902-540-8359  
Sequence 8359, Application US/09902540

```
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 8359
; LENGTH: 444
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-8359
```

```
Query Match 71.7%; Score 17.2; DB 4; Length 444;
Best Local Similarity 66.7%; Pred. No. 5.4e+02;
Matches 16; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 GTCTGTGGCGCGTGTACGCGCTC 24
DB 270 GCCCTGTGGCGCGTGTACGCGCTC 293
```

```
RESULT 30
US-09-252-991A--6007/C
; Sequence 6007, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6007
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6007
```

```
Query Match 71.7%; Score 17.2; DB 4; Length 585;
Best Local Similarity 66.7%; Pred. No. 5.4e+02;
Matches 16; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 GTCTGTGGCGCGTGTACGCGSAC 24
DB 552 GCCGTCTGTGGCGGTGTACCGGAC 529
```

Search completed: July 20, 2005, 16:48:37  
Job time : 131.714 secs





C 98	17.4	72.5	553	2	AM498322	AM498322	660046B11	C 171	17.2	71.7	419	7	CO525050	CO525050	3530_1_16
C 99	17.4	72.5	560	4	CE731010	CE731010	l1gr-g8s-	C 172	17.2	71.7	423	7	CK930464	CK930464	p5imgc_00
C 100	17.4	72.5	569	4	BI340535	BI340535	365832 MA	C 173	17.2	71.7	426	7	CK930385	CK930385	p5imgc_00
C 101	17.4	72.5	578	4	BI340506	BI340506	365796 MA	C 174	17.2	71.7	432	7	CK930422	CK930422	p5imgc_00
C 102	17.4	72.5	620	8	BZ708597	BZ708597	OGBAV35TM	C 175	17.2	71.7	434	2	AM374956	MR0-CT006	AM374956
C 103	17.4	72.5	643	7	CO521649	CO521649	3530_1_14	C 176	17.2	71.7	436	7	BB671056	BB671056	p5imgc_00
C 104	17.4	72.5	657	7	CO523848	CO523848	3530_1_15	C 177	17.2	71.7	446	7	CK930114	CK930114	p5imgc_00
C 105	17.4	72.5	680	8	BZ564040	BZ564040	pac62-164	C 178	17.2	71.7	450	7	CK930187	CK930187	p5imgc_00
C 106	17.4	72.5	699	8	AY104786	AY104786	z6a mayr	C 179	17.2	71.7	450	7	CK930428	CK930428	p5imgc_00
C 107	17.4	72.5	721	8	BZ708588	BZ708588	OGBAV35TC	C 180	17.2	71.7	451	7	CK930640	CK930640	p5imgc_00
C 108	17.4	72.5	791	4	CG372496	CG372496	PUNSR04TD	C 181	17.2	71.7	454	7	CK929187	CK929187	p5imgc_00
C 109	17.4	72.5	821	4	BC320069	BC320069	Zm03_01f0	C 182	17.2	71.7	459	4	B1795700	B1795700	p5imgc_00
C 110	17.4	72.5	835	6	CA085587	CA085587	SCUTAM209	C 183	17.2	71.7	460	7	CK929968	CK929968	p5imgc_00
C 111	17.4	72.5	874	2	BF215026	BF215026	60184735B	C 184	17.2	71.7	461	7	CNI181198	CNI181198	SGP169f12
C 112	17.4	72.5	1028	9	CG115513	CG115513	PUI065TB	C 185	17.2	71.7	463	7	CK930564	CK930564	p5imgc_00
C 113	17.4	72.5	1064	9	CNS02073	AL116666	Tetradon	C 186	17.2	71.7	466	7	CK929233	CK929233	p5imgc_00
C 114	17.4	72.5	1347	3	BC030024	BC030024	Homo sapi	C 187	17.2	71.7	466	7	CK930176	CK930176	p5imgc_00
C 115	17.2	71.7	13547	2	BB602482	BB602482	BB602482	C 188	17.2	71.7	469	7	CK930176	CK930176	p5imgc_00
C 116	17.2	71.7	206	8	BH256468	BH256468	KG03203 D	C 189	17.2	71.7	472	5	BX526676	BX526676	p5imgc_00
C 117	17.2	71.7	208	8	BH538521	BH538521	BB38521	C 190	17.2	71.7	474	1	AL830986	AL830986	p5imgc_00
C 118	17.2	71.7	218	7	CK930931	CK930931	p5imgc_00	C 191	17.2	71.7	475	7	CK929207	CK929207	p5imgc_00
C 119	17.2	71.7	220	7	CK930073	CK930073	p5imgc_00	C 192	17.2	71.7	476	7	CK930749	CK930749	p5imgc_00
C 120	17.2	71.7	239	7	CK929744	CK929744	p5imgc_00	C 193	17.2	71.7	476	7	CO524912	CO524912	3530_1_16
C 121	17.2	71.7	239	7	CK931038	CK931038	p5imgc_00	C 194	17.2	71.7	479	5	BR459026	BR459026	p5imgc_00
C 122	17.2	71.7	244	7	CK930396	CK930396	p5imgc_00	C 195	17.2	71.7	479	7	CK928916	CK928916	p5imgc_00
C 123	17.2	71.7	244	7	CK930775	CK930775	p5imgc_00	C 196	17.2	71.7	480	7	CK929740	CK929740	p5imgc_00
C 124	17.2	71.7	251	7	CK930934	CK930934	p5imgc_00	C 197	17.2	71.7	480	7	CK930518	CK930518	p5imgc_00
C 125	17.2	71.7	260	7	CK929060	CK929060	p5imgc_00	C 198	17.2	71.7	481	9	AG257747	Lotus cor	AG257747
C 126	17.2	71.7	267	7	CK930516	CK930516	p5imgc_00	C 199	17.2	71.7	484	7	CK929216	CK929216	p5imgc_00
C 127	17.2	71.7	276	7	H27643	H27643	Y1E1h01.T1	C 200	17.2	71.7	485	6	BY597533	BY597533	p5imgc_00
C 128	17.2	71.7	277	7	CK930449	CK930449	p5imgc_00	C 201	17.2	71.7	486	7	CK930043	CK930043	p5imgc_00
C 129	17.2	71.7	282	7	CK929198	CK929198	p5imgc_00	C 202	17.2	71.7	487	7	CK929590	CK929590	p5imgc_00
C 130	17.2	71.7	286	7	CK930413	CK930413	p5imgc_00	C 203	17.2	71.7	490	7	CK929527	CK929527	p5imgc_00
C 131	17.2	71.7	292	2	BB208092	BB208092	BB208092	C 204	17.2	71.7	491	7	CK131411	CK131411	LP02583.3
C 132	17.2	71.7	297	2	BB360214	BB360214	BB360214	C 205	17.2	71.7	491	7	CK929837	CK929837	p5imgc_00
C 133	17.2	71.7	298	7	CK929949	CK929949	p5imgc_00	C 206	17.2	71.7	494	7	CK930352	CK930352	p5imgc_00
C 134	17.2	71.7	299	7	CK930369	CK930369	p5imgc_00	C 207	17.2	71.7	494	7	CK929494	CK929494	p5imgc_00
C 135	17.2	71.7	300	7	CK929576	CK929576	p5imgc_00	C 208	17.2	71.7	497	7	CK929601	CK929601	p5imgc_00
C 136	17.2	71.7	304	7	CK929850	CK929850	p5imgc_00	C 209	17.2	71.7	497	7	CK930832	CK930832	p5imgc_00
C 137	17.2	71.7	305	7	CO524947	CO524947	3530_1_16	C 210	17.2	71.7	500	6	CA130140	CA130140	SCVPLR201
C 138	17.2	71.7	306	7	CP857349	CP857349	p5imgc_00	C 211	17.2	71.7	500	7	CK929718	CK929718	p5imgc_00
C 139	17.2	71.7	308	7	CP853616	CP853616	p5imgc_00	C 212	17.2	71.7	505	7	CK929999	CK929999	p5imgc_00
C 140	17.2	71.7	309	1	AJ603499	AJ603499	p5imgc_00	C 213	17.2	71.7	505	7	CK929574	CK929574	p5imgc_00
C 141	17.2	71.7	320	7	CK930307	CK930307	p5imgc_00	C 214	17.2	71.7	507	7	CK929440	CK929440	p5imgc_00
C 142	17.2	71.7	322	9	CE553081	CE553081	l1gr-g8s-	C 215	17.2	71.7	507	7	CK929523	CK929523	p5imgc_00
C 143	17.2	71.7	328	7	CK929725	CK929725	p5imgc_00	C 216	17.2	71.7	507	7	CK930438	CK930438	p5imgc_00
C 144	17.2	71.7	337	2	BR335934	BR335934	CM4-CT068	C 217	17.2	71.7	513	1	AV430005	AV430005	p5imgc_00
C 145	17.2	71.7	341	2	AM579227	AM579227	PM2-DT004	C 218	17.2	71.7	517	7	CK929585	CK929585	p5imgc_00
C 146	17.2	71.7	349	7	CK929878	CK929878	p5imgc_00	C 219	17.2	71.7	519	7	CK930282	CK930282	p5imgc_00
C 147	17.2	71.7	355	6	BY594589	BY594589	BY594589	C 220	17.2	71.7	521	7	CK928951	CK928951	p5imgc_00
C 148	17.2	71.7	356	7	CK929008	CK929008	p5imgc_00	C 221	17.2	71.7	524	2	AM564100	AM564100	LG1_282.E
C 149	17.2	71.7	359	8	BZ293509	BZ293509	CG0368_r1	C 222	17.2	71.7	524	7	CK929640	CK929640	p5imgc_00
C 150	17.2	71.7	365	5	BP087529	BP087529	p5imgc_00	C 223	17.2	71.7	526	7	CK930103	CK930103	p5imgc_00
C 151	17.2	71.7	372	7	CK929358	CK929358	p5imgc_00	C 224	17.2	71.7	527	7	CK928894	CK928894	p5imgc_00
C 152	17.2	71.7	381	7	CK929756	CK929756	p5imgc_00	C 225	17.2	71.7	529	7	CK930722	CK930722	p5imgc_00
C 153	17.2	71.7	383	7	CO191444	CO191444	BC030874.5	C 226	17.2	71.7	530	7	CK930864	CK930864	p5imgc_00
C 154	17.2	71.7	388	2	AM659731	AM659731	97381 MAR	C 227	17.2	71.7	532	2	BB468773	BB468773	IPdKf004
C 155	17.2	71.7	390	7	CK929340	CK929340	p5imgc_00	C 228	17.2	71.7	532	7	CK929810	CK929810	p5imgc_00
C 156	17.2	71.7	390	7	U47733	U47733	HSU47733_Hu	C 229	17.2	71.7	535	1	A1261133	A1261133	LP05241.3
C 157	17.2	71.7	396	5	BP086909	BP086909	p5imgc_00	C 230	17.2	71.7	538	7	CK931028	CK931028	p5imgc_00
C 158	17.2	71.7	396	5	CE158344	CE158344	l1gr-g8s-	C 231	17.2	71.7	539	1	A1258473	A1258473	LP016807.5
C 159	17.2	71.7	397	7	CK928666	CK928666	p5imgc_00	C 232	17.2	71.7	539	1	AL915020	AL915020	p5imgc_00
C 160	17.2	71.7	399	7	CK929789	CK929789	p5imgc_00	C 233	17.2	71.7	539	7	CK656943	CK656943	LP22811.3
C 161	17.2	71.7	401	7	CP856066	CP856066	p5imgc_00	C 234	17.2	71.7	539	7	CK929836	CK929836	p5imgc_00
C 162	17.2	71.7	402	7	CK930024	CK930024	p5imgc_00	C 235	17.2	71.7	540	7	CK930378	CK930378	p5imgc_00
C 163	17.2	71.7	403	7	CK929466	CK929466	p5imgc_00	C 236	17.2	71.7	541	7	CK928949	CK928949	p5imgc_00
C 164	17.2	71.7	404	5	BY286951	BY286951	BY286951	C 237	17.2	71.7	543	1	BE419770	BE419770	WMS017.C1
C 165	17.2	71.7	409	7	CK929501	CK929501	p5imgc_00	C 238	17.2	71.7	544	2	AV626769	AV626769	p5imgc_00
C 166	17.2	71.7	409	7	CK930143	CK930143	p5imgc_00	C 239	17.2	71.7	544	7	CK929681	CK929681	p5imgc_00
C 167	17.2	71.7	413	7	CK930476	CK930476	p5imgc_00	C 240	17.2	71.7	548	7	CK930548	CK930548	p5imgc_00
C 168	17.2	71.7	413	7	CK930827	CK930827	p5imgc_00	C 241	17.2	71.7	548	2	BR013997	BR013997	p5imgc_00
C 169	17.2	71.7	417	7	CK930196	CK930196	p5imgc_00	C 242	17.2	71.7	548	7	CK930094	CK930094	p5imgc_00
C 170	17.2	71.7	419	7	CK930313	CK930313	p5imgc_00	C 243	17.2	71.7	549	4	BI344979	BI344979	373645_MA

C 244	17.2	71.7	549	7	CK930091	CK930091 p5imgc_00
C 245	17.2	71.7	555	7	CK929948	CK929948 p5imgc_00
C 246	17.2	71.7	555	7	CK930699	CK930699 p5imgc_00
C 247	17.2	71.7	556	7	CK929021	CK929021 p5imgc_00
C 248	17.2	71.7	557	6	CD488880	CD488880 T15_F07_T
C 249	17.2	71.7	562	7	CK929425	CK929425 p5imgc_00
C 250	17.2	71.7	563	7	CK930008	CK930008 p5imgc_00
C 251	17.2	71.7	567	7	CK929283	CK929283 p5imgc_00
C 252	17.2	71.7	570	7	CK930365	CK930365 p5imgc_00
C 253	17.2	71.7	571	7	CF181700	CF181700 818615 MA
C 254	17.2	71.7	572	7	CK929566	CK929566 p5imgc_00
C 255	17.2	71.7	573	1	AU289764	AU289764 AU289764
C 256	17.2	71.7	574	6	CA287868	CA287868 SCSFSD206
C 257	17.2	71.7	575	7	CK929325	CK929325 p5imgc_00
C 258	17.2	71.7	576	7	CK929866	CK929866 p5imgc_00
C 259	17.2	71.7	577	7	CK930079	CK930079 p5imgc_00
C 260	17.2	71.7	579	7	CK929895	CK929895 p5imgc_00
C 261	17.2	71.7	579	9	CE411108	CE411108 ligf-gfs-
C 262	17.2	71.7	581	6	CD810583	CD810583 1336 Ling
C 263	17.2	71.7	582	4	BM076672	BM076672 T1BST-A08
C 264	17.2	71.7	582	7	CK930208	CK930208 p5imgc_00
C 265	17.2	71.7	585	7	CK931090	CK931090 p5imgc_00
C 266	17.2	71.7	587	7	CF850440	CF850440 pSMA014xf
C 267	17.2	71.7	587	7	CK929840	CK929840 p5imgc_00
C 268	17.2	71.7	587	7	CK929840	CK929840 p5imgc_00
C 269	17.2	71.7	589	5	BQ464452	BQ464452 HF02F05r
C 270	17.2	71.7	589	5	AG271420	AG271420 Cyanidlos
C 271	17.2	71.7	594	7	CK929692	CK929692 p5imgc_00
C 272	17.2	71.7	594	4	BG368479	BG368479 HVSME1001
C 273	17.2	71.7	595	7	CK929242	CK929242 p5imgc_00
C 274	17.2	71.7	598	7	CK929860	CK929860 p5imgc_00
C 275	17.2	71.7	601	7	CK930633	CK930633 p5imgc_00
C 276	17.2	71.7	602	7	CK929936	CK929936 p5imgc_00
C 277	17.2	71.7	604	7	CK930067	CK930067 p5imgc_00
C 278	17.2	71.7	611	7	CK930981	CK930981 p5imgc_00
C 279	17.2	71.7	625	7	CK930981	CK930981 p5imgc_00
C 280	17.2	71.7	629	8	BH840371	BH840371 LMC20001
C 281	17.2	71.7	630	4	BI724091	BI724091 1031069A1
C 282	17.2	71.7	630	4	CK930334	CK930334 p5imgc_00
C 283	17.2	71.7	632	9	CG205632	CG205632 PULCS01TD
C 284	17.2	71.7	634	9	CG840062	CG840062 Ynhw0676
C 285	17.2	71.7	636	4	BI548787	BI548787 603189188
C 286	17.2	71.7	640	4	BG518790	BG518790 602578382
C 287	17.2	71.7	642	4	BJ649918	BJ649918 BJ649918
C 288	17.2	71.7	645	7	CK930223	CK930223 p5imgc_00
C 289	17.2	71.7	645	8	BZ329818	BZ329818 hv88b05_b
C 290	17.2	71.7	654	5	BQ812592	BQ812592 1030031B0
C 291	17.2	71.7	667	7	CK930145	CK930145 SB0201B2
C 292	17.2	71.7	677	7	CK940602	CK940602 4064046_B
C 293	17.2	71.7	680	9	CG242034	CG242034 OGV248TH
C 294	17.2	71.7	684	7	CK930290	CK930290 p5imgc_00
C 295	17.2	71.7	695	6	CD924088	CD924088 G750_111h
C 296	17.2	71.7	703	5	BU452052	BU452052 603218379
C 297	17.2	71.7	713	5	BU213716	BU213716 603756629
C 298	17.2	71.7	714	7	CV104155	CV104155 AGENCOURT
C 299	17.2	71.7	716	7	CK940880	CK940880 4064430_B
C 300	17.2	71.7	719	5	BU137350	BU137350 603007751

## ALIGNMENTS

RESULT 1  
BE869584/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BE869584  
BE869584  
BE869584.1  
EST.  
Homo sapiens  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

1059 bp  
mRNA  
linear  
EST 20-OCT-2000  
60144569661 NIH\_MGC\_65 Homo sapiens cDNA clone IMAGE:3849863 5',  
mRNA sequence.  
GI:10318464

REFERENCE  
1. (bases 1 to 1059)  
NIH-MGC <http://mgi.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)

FEATURES  
SOURCE  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: LAM9567 row: p column: 24  
High quality sequence stop: 719.  
Location/Qualifiers  
1. 1059  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3849863"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_65"  
/note="Organ: colon; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.8 kb. Library constructed by Life Technologies."

## ORIGIN

Query Match 80.8%; Score 19.4; DB 2; Length 1059;  
Best Local Similarity 73.9%; Pred. No. 1.5e+03;  
Matches 17; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCTGTGGCCGCGTACCGSAC 24  
DB 967 TCTGTGGCCGCGTACCGTAC 945

RESULT 2  
CN719278/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

CN719278  
E0781F12-5 NIA Mouse four-cell-Embryo cDNA library (long) Mus  
musculus cDNA clone NIA:E0781F12 IMAGE:30903815 5', mRNA sequence.  
CN719278  
CN719278.1 GI:47488663  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euteleostomi; Rodentia; Sciurognathia; Muridae; Murine; Mus.  
1 (bases 1 to 132)  
Sharov,A.A., Piao,Y., Matoba,R., Dudekula,D.B., Qian,Y.,  
VanBuren,V., Falco,G., Martin,P.R., Steagy,C.A., Bassey,U.C.,  
Mang,Y., Carter,M.G., Hamatani,T., Albe,K., Akutsu,H., Sharova,L.,  
Tanaka,T.S., Kimber,W.L., Yoshikawa,T., Jaradat,S.A., Pantano,S.,  
Nagaraja,R., Boheler,K.R., Taub,D., Hodess,R.J., Longo,D.L.,  
Schlesinger,D., Keller,U., Klotz,E., Kelsoe,G., Umezawa,A.,  
Vescovi,A.L., Rossant,J., Kunath,T., Hogan,B.L., Curci,A.,  
D'Urso,M., Klesio,U., Hide,W. and Ko,M.S.  
Transcriptome analysis of mouse stem cells and early embryos  
PLoS Biol. 1 (3): 410-419 (2003)  
Contact: Dawood B. Dudekula  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: [cdnaelgaun-grc.nia.nih.gov](mailto:cdnaelgaun-grc.nia.nih.gov)  
Plate: E0781 row: F column: 12  
Seq primer: M13 Reverse  
High quality sequence stop: 132  
POLYA=No.  
Location/Qualifiers

## FEATURES

## SOURCE

1. 132  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /db\_xref="taxon:10090"  
 /clone="NIA:E0781P12 IMAGE:30903815"  
 /tissue\_type="4-cell stage embryo"  
 /dev\_stage="4-cell"  
 /lab\_host="DH10B"  
 /clone\_lib="NIA Mouse four-cell-Embryo cDNA library (Long)"  
 /note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (<http://lgsun.grc.nia.nih.gov/cDNA>). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMD: 11544199]). The mRNAs were extracted from a pool of 360 embryos at 4-cell stage. Double-stranded cDNAs were synthesized with an oligo(dT) primer (Invitrogen):  
 5'-pGACTAGTCTAGTCGCGAGCGCGCCCTTTTCTTTT-3' from 10.8ng of mRNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lone-Linker L1-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Tag polymerase (Takara) with a primer SalI-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2Kb. The library was constructed by Yulan Piao."

## ORIGIN

Query Match 78.3%; Score 18.8; DB 7; Length 132;  
 Best Local Similarity 70.8%; Pred. No. 2.8e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCTGTGTCGCGCGGTCGCGGAC 24  
 |||||:||||:||||:||||:  
 Db 86 GTCTGTGTCGCGCGGTCGCGGAC 63

## RESULT 3

BB869974/c

LOCUS BB869974 RIKEN full-length enriched, 13 days embryo spinal cord Mus 247 bp mRNA linear EST 27-NOV-2001  
 DEFINITION musculus cDNA clone G630017E12 5', mRNA sequence.

ACCESSION BB869974

VERSION BB869974.1 GI:17116184

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 247)

REFERENCE

AUTHORS

Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hasegaki,T., Hayatsu,N., Hiramoto,K., Hirata,K., Hirozane,T., Imocani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Kondo,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numaaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Saeki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,T., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)

TITLE

JOURNAL

COMMENT Unpublished (2001)

Laboratory for Genome Exploration Research Group, RIKEN Genomic

## FEATURES

source

Sciences Center(GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.jp, URL:<http://genome.gsc.riken.jp/>  
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Irawa,M., Ohara,E., Watanhiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multichipillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
 Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.  
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.  
 e mouse tissues.

## ORIGIN

Query Match 78.3%; Score 18.8; DB 2; Length 247;  
 Best Local Similarity 70.8%; Pred. No. 2.8e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCTGTGTCGCGCGGTCGCGGAC 24  
 |||||:||||:||||:||||:  
 Db 100 GTCTGTGTCGCGCGGTCGCGGAC 77

## RESULT 4

CK929651/c

LOCUS CK929651 315 bp mRNA linear EST 22-APR-2004  
 DEFINITION p57mgc\_001191 Normalized Magnaporthe grisea cDNA p57mgc\_001191, mRNA sequence.

ACCESSION CK929651

VERSION CK929651.1 GI:45417502

KEYWORDS EST.

SOURCE

ORGANISM

Magnaporthe grisea (anamorph: Pyricularia grisea)  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetes; Insectae sedis; Magnaportheaceae; Magnaporthe.  
 1 (bases 1 to 315)  
 Chen,B., Li,Y., Peng,Y., Dong,H. and Li,D.  
 Large-scale identification of ESTs from Magnaporthe grisea by normalized cDNA library sequencing  
 Unpublished (2004)

JOURNAL

COMMENT

Contact: Baozhan Chen, Youzhi Li  
 Laboratory of Subtropical Bioresource Conservation and Utilization  
 Guangxi University, China Agricultural University, Zhejiang  
 University  
 Daxue Road 100#, Nanning, Guangxi, 530004, China  
 Tel: 0086-771-3237873  
 Fax: 0086-771-3237873  
 Email: chenbs@n.gx.cninfo.net, URL: <http://www.estarray.org>

Seq primer: M13 forward primer.  
Location/Qualifiers  
1. .315  
/organism="Magnaporthe grisea"  
/mol\_type="mRNA"  
/db\_xref="taxon:148305"  
/clone="p5rmgc.001191"  
/issue\_type="Myceillum, conidium, germinating conidium, swelling appressorium, mature appressorium, penetration peg"  
/dev\_stage="Myceillum, conidium, germinating conidium, swelling appressorium, mature appressorium, penetration peg"  
/clone\_lib="Normalized Magnaporthe grisea cDNA pGEM-T Easy library"  
/note="Vector: pGEM-T Easy"

ORIGIN  
Query Match 78.3%; Score 18.8; DB 7; Length 315;  
Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCCGTGCGCGCGTGAACCGGAC 24  
Db 193 GTCCGTGCGCGCGTGAACCGGAC 170

RESULT 5  
BB868549/c  
LOCUS  
DEFINITION BB868549 RIKEN full-length enriched, 16 days neonate male diencephalon Mus musculus cDNA clone G63006D18 5', mRNA sequence.  
BB868549  
VERSION BB868549.1 GI:17114759  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 338)  
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hangaki,T., Hayatsu,N., Hiramoto,K., Hirakawa,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasabe,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watanishi,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (AKimura,T., et al. 2001)

TITLE Unpublished (2001)  
JOURNAL Contact: Yoshihide Hayashizaki  
COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)  
Wagii,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanishi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)  
Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,

Sugahara,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.  
e mouse tissues.  
Location/Qualifiers  
1. .338  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="G63006D18"  
/sex="male"  
/issue\_type="dienecephalon"  
/dev\_stage="16 days neonate"  
/clone\_lib="RIKEN full-length enriched, 16 days neonate male dienecephalon"

ORIGIN  
Query Match 78.3%; Score 18.8; DB 2; Length 338;  
Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCCGTGCGCGCGTGAACCGGAC 24  
Db 123 GTCCGTGCGCGCGCGTGAACCGGAC 100

RESULT 6  
BY280059/c  
LOCUS  
DEFINITION BY280059 RIKEN full-length enriched, visual cortex Mus musculus cDNA clone K430344N19 5', mRNA sequence.  
BY280059  
VERSION BY280059.1 GI:26470396  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 396)  
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaide,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Wagii,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schombach,C., Gojodori,T., Balderelli,R., Hill,D.P., Bult,C., Hume,D.A., Quakenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Brad,D., Brusic,V., Chotha,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochik,I.V., Lee,Y., Lennard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Nomura,K., Okido,T., Pavan,W.J., Perlea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,U.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Sempile,C.A., Seton,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Varado,R., Wagner,L., Wahlstedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wymshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Kono,H., Nakamura,M., Sakazume,N., Sato,K., Shibata,T., Waki,K., Kawai,J., Aizawa,K., Akakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,U., Birney,B. and Hayashizaki,Y.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)

TITLE  
JOURNAL

MEDLINE  
22354683  
12466851  
COMMENT  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/

Alizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S.,  
Hirozane, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Kono, H.,  
Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Numazaki, R.,  
Ohno, M., Sakai, K., Sakazume, N., Sasaki, D., Sato, K., Shibata, K.,  
Shiraki, T., Tagami, M., Waki, K., Wataniki, A., Muramatsu, M. and  
Hayashizaki, Y. Direct Submission  
Computational Analysis of Full-length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1577-1771 (2000)  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in Riken.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Tissues were provided by Michela Fagioli and Takao K. Hensch ( )  
Laboratory for Neuronal Circuit Development Brain Science Institute  
RIKEN 2-1 Hirsawa, Wako-shi, Saitama 351-0198 Japan.) whose  
assistance we gratefully acknowledge. Please visit our web site  
(http://genome.gsc.riken.go.jp) for further details.

FEATURES  
source  
1..396  
Location/Qualifiers

/organism="Mus musculus"  
/mol\_type="mRNA"  
/accession="U57BL/6J"  
/db\_xref="taxon:10090"  
/clone="K430344N19"  
/issue\_type="Visual cortex"  
/clone\_lib="RIKEN full-length enriched, visual cortex"

## ORIGIN

Query Match 78.3%; Score 18.8; DB 5; Length 396;  
Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCCGTGTCGCGCGTACGCGSAC 24  
|||||:|||||:|||||:  
Db 98 GTCCGTGTCGCGCGGAGCGCGAC 75

RESULT 7  
LOCUS CK929166/c 428 bp mRNA linear EST 22-APR-2004  
DEFINITION p5imgc\_000411 Normalized Magnaporthe grisea cDNA pGEM-T Easy  
library Magnaporthe grisea cDNA clone p5imgc\_000411, mRNA sequence.  
ACCESSION CK929166  
VERSION CK929166.1 GI:45417017  
KEYWORDS EST.  
SOURCE Magnaporthe grisea (anamorph: Pyricularia grisea)  
ORGANISM Magnaporthe grisea  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.  
1 (bases 1 to 428)  
Chen, B., Li, Y., Peng, Y., Dong, H. and Li, D.  
Large-scale identification of ESTs from Magnaporthe grisea by  
normalized cDNA library sequencing

JOURNAL  
COMMENT

Unpublished (2004)  
Contact: Baoshan Chen, Youzhi Li  
Laboratory of Subtropical Bioresource Conservation and Utilization  
Guangxi University, China Agricultural University, Zhenjiang  
University  
Daxue Road 100#, Nanning, Guangxi, 530004, China  
Tel: 0086-771-3339566  
Fax: 0086-771-3337873  
Email: chenbs@n.gxu.cninfo.net, URL: http://www.estarray.org  
Seq primer: M13 forward primer.

FEATURES  
source

1..428  
Location/Qualifiers  
/organism="Magnaporthe grisea"  
/mol\_type="mRNA"  
/db\_xref="taxon:148305"  
/clone="p5imgc\_000411"  
/issue\_type="Myceallium, conidium, germinating conidium,  
swelling appressorium, mature appressorium, penetration  
peg"  
/dev\_stage="Myceallium, conidium, germinating conidium,  
swelling appressorium, mature appressorium, penetration  
peg"  
/clone\_lib="Normalized Magnaporthe grisea cDNA pGEM-T Easy  
library"  
/note="Vector: pGEM-T Easy"

## ORIGIN

Query Match 78.3%; Score 18.8; DB 7; Length 428;  
Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCCGTGTCGCGCGTACGCGSAC 24  
|||||:|||||:|||||:  
Db 181 GTCCGTGTCGCGCGGAGCGCGAC 158

RESULT 8  
LOCUS BB863614/c

DEFINITION BB863614 RIKEN full-length enriched, pooled cell lines,  
RCB-0544, etc. Mus musculus cDNA clone G430038G07 5', mRNA sequence.  
ACCESSION BB863614  
VERSION BB863614.1 GI:17109824  
KEYWORDS EST.

## SOURCE

Mus musculus  
(house mouse)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS

1 (bases 1 to 432)  
Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,  
Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Imotani, K.,  
Ishii, Y., Ito, M., Kawai, J., Kojima, Y., Kono, H., Kouda, M.,  
Matsuyama, T., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R.,  
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sakazume, N.,  
Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Suzuki, H., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Tanaka, T., Tomaru, A., Toya, T., Wataniki, A., Yasumishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.  
2001)

## TITLE

JOURNAL  
COMMENT

Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi.K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
 Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,  
 Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A.  
 and Hayashizaki,Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,  
 Sugahara,Y. and Hayashizaki,Y.  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
 further details.  
 e mouse tissues.

#### FEATURES

##### source

1. 432  
 location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="G430038G07"  
 /clone.lib="RIKEN full-length enriched, pooled cell lines,  
 RCB-0544, etc."  
 /note="pooled tissues: (tissue\_type=bladder,sex=mix,  
 cell\_line=RCB-0544 MBT-2), (tissue\_type=embryo,sex=mix,  
 cell\_line=RCB-0549 Cle-H3), (strain=DDD,tissue\_type=spleen,  
 sex=mix,cell\_line=RCB-0559  
 K-1.fl), (strain=BALB/C,tissue\_type=blood,sex=mix,  
 cell\_line=RCB-0035 WEHI-3), (strain=C57BL/6,  
 tissue\_type=skin,sex=mix,cell\_line=RCB-1283 B16  
 melanoma), (strain=BALB/C,tissue\_type=sarcoma,sex=mix,  
 cell\_line=RCB-0464 Meth-A), (strain=ICR,tissue\_type=embryo,  
 sex=mix,cell\_line=RCB-0545 OHTA), (strain=C3H/2J,dev\_stage=7  
 days embryo,sex=mix,cell\_type=multipotent stem  
 cell,cell\_line=CRL-2070  
 NE), (strain=BALB/C,tissue\_type=kidney,sex=mix,  
 cell\_line=CCL-142 RAG), (strain=Swiss  
 Webster,tissue\_type=submandibular  
 gland,sex=male,cell\_line=CRL-1734 SCA-9 clone  
 15), (strain=BALB/CROs,dev\_stage=1.5  
 years,tissue\_type=mammary  
 gland,sex=female,cell\_line=CRL-2116  
 JC), (strain=BALB/C,sex=mix,cell\_line=CRL-1751 WEHI  
 164), (strain=C3H,tissue\_type=brain,sex=mix,  
 cell\_line=CRL-1443 BC3H1), (strain=BALB/C,tissue\_type=B  
 lymphocyte,sex=mix,cell\_type=B cells,cell\_line=CRL-1669  
 BCL1 Clone 13.20-3B3), (strain=BDX,dev\_stage=5  
 months,tissue\_type=bone marrow,sex=female,cell\_type=stroma  
 cell,cell\_line=CRL-2028  
 SR-4987), (strain=C57BL/6,tissue\_type=testis,  
 sex=mix,cell\_type=leydig cells,cell\_line=CRL-2065  
 MUTC-1), (strain=BALB/C X NZB F1,tissue\_type=B  
 lymphocyte,sex=mix,cell\_type=B cells,cell\_line=CRL-1702  
 WEHI 231)"

#### ORIGIN

Query Match 78.3% Score 18.8; DB 2; Length 432;  
 Best Local Similarity 70.8% Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 GTCCGTGCGCCGTCGACGCGAC 24  
 |||||:|||||:|||||:|||||:|||||:  
 Db 101 GTCTGGGGGCGCGGACGCGCAC 78

RESULT 9  
 BB839789/c 447 bp mRNA linear EST 21-NOV-2001  
 LOCUS BB839789 RIKEN full-length enriched, 12 days embryo whole body Mus  
 DEFINITION BB839789 musculus cDNA clone E970001N24 5', mRNA sequence.  
 ACCESSION BB839789  
 VERSION BB839789.1 GI:17040520

#### KEYWORDS

EST.  
 Mus musculus (house mouse)

#### SOURCE

Mus musculus

#### ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

#### REFERENCE

1 (bases 1 to 447)

#### AUTHORS

Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hangaki,T.,  
 Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K.,  
 Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Kono,H., Kouda,M.,  
 Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,  
 Okazaki,T., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,  
 Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,  
 Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,  
 Tanaka,T., Tomaru,A., Toyota,T., Watabiki,A., Yasunishi,A.,  
 Muramatsu,M. and Hayashizaki,Y.  
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.  
 2001)

#### TITLE

#### JOURNAL

Unpublished (2001)  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216

#### COMMENT

Email: genome-res@gsr.riken.jp, URL: <http://genome.gsc.riken.jp/>  
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,  
 Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi.K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
 Watabiki,M., Yoneda,Y., Ishikawa,T., Oawa,K., Tanaka,T.,  
 Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A.  
 and Hayashizaki,Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,  
 Sugahara,Y. and Hayashizaki,Y.  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
 further details.  
 e mouse tissues.

#### FEATURES

##### source

1. 447  
 location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="E970001N24"  
 /tissue\_type="whole body"  
 /dev\_stage="12 days embryo"  
 /clone.lib="RIKEN full-length enriched, 12 days embryo  
 whole body"

#### ORIGIN

Query Match 78.3% Score 18.8; DB 2; Length 447;  
 Best Local Similarity 70.8% Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 GTCCGTGCGCCGTCGACGCGAC 24  
 |||||:|||||:|||||:|||||:|||||:  
 Db 73 GTCTGGGGGCGCGGACGCGCAC 50

RESULT 10  
 AA036131 481 bp mRNA linear EST 26-AUG-1996  
 LOCUS AA036131  
 DEFINITION AA036131 Soares mouse embryo NDBE13.5 14.5 Mus musculus cDNA  
 clone IMAGE:468931 5', mRNA sequence.



```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:468931"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="MDH10B"
/clone_1ib="Soares mouse embryo NBME13.5 14.5"
/note="vector: pYT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer 15',
TGTTACCAATCTGAAGTCGAGCGCGCGGAAATTTTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
Staele Univ., from 2 j], double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pYT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo. "

```

REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 510)			
TITLE	Piao, Y., Dudekula, D.B., Qian, Y., Martin, P. R., Alba, K., Vescovi, A. L. and Ko, M. S. H.			
JOURNAL	Systematic Analyses of NIA Mouse Neural Stem Cell			
COMMENT	(Undifferentiated) cDNA Library (Long)			
	Unpublished (2002)			
	Contact: Dawood B. Dudekula			
	Laboratory of Genetics			
	National Institute on Aging/National Institutes of Health			
	333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA			
	Email: cdna@lgsun.grc.nia.nih.gov			
	Plate: K0972 row: B column: 03			
	Seq primer: -21M13 Reverse			
	High quality sequence stop: 510			
	POLYA=No.			
FEATURES	Location/Qualifiers			
SOURCE	1. 510			
	/organism="Mus musculus"			
	/mol_type="mRNA"			
	/strain="CD1"			
	/db_xref="taxon:10090"			
	/db_xref="taxon:10090"			
	/clone="NIA:K0972B03 IMAGE:30091118"			
	/dev_stage="Adult"			
	/lab_host="DHI08"			
	/clone_1lb="NIA Mouse Neural Stem Cell (Undifferentiated)			
	cDNA library (Long)"			
	/note="Vector: pSPORT1 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were obtained from Dr. Angelo L. Vescovi (Institute for Stem Cell Research, Italy). Double-stranded cDNAs were synthesized with an oligo(dT) primer [Invitrogen: 5'-pGACTGCTTCAGATCGGAGCGCCGCCCTTTTCTTTT-3'] from 2.0 Microgram of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker ltr-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DHI08 E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 3.8 kb. The library was constructed by Yulan Piao."			
ORIGIN				
Query Match	78.3%;	Score 18.8;	DB 6;	Length 510;
Best Local Similarity	70.8%;	Pred. No. 2.7e+03;		
Matches 17; Conservative	5;	Mismatches 2;	Indels 0;	Gaps 0;
Cy	1 GTCCTGAGTGGCGTSGACGCGAC 24			
Db	106 GTCTTGAGGCGCGGAGCGCGAC 83			
RESULT 12				
CB606070/c	516 bp mRNA linear EST 16-MAY-2003			
LOCUS				
DEFINITION	AGNNNNC:NRHY7-00021-A10-A nrhy7 (10850) Rattus norvegicus cDNA			
ACCESSION	CB606070			
VERSION	CB606070.1 GI:29545662			
KEYWORDS	EST.			
SOURCE	Rattus norvegicus (Norway rat)			
ORGANISM	Rattus norvegicus			



Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 1 (bases 1 to 516)  
 Amgen EST Program.  
 Amgen Rat EST Program  
 Unpublished (2003)  
 Contact: Dan Fitzpatrick  
 Amgen, Inc  
 One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA  
 Tel: 805 447-4881  
 Plate: 00021 row: a column: 10.  
 Location/Qualifiers  
 1..516  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10116"  
 /clone="nrhy7-00021-a10"  
 /clone\_1lb="nrhy7 (10850)"  
 /note="Vector: pSPOR1; Site 1: SalI; Site 2: NotI; W Rat hypothalamus adult female Wistar rat avg. insert size 2.3 kb fraction 6 and 7"

ORIGIN  
 Query Match 78.3%; Score 18.8; DB 6; Length 516;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Oy 1 GTCTGTGCGCGCGTACGCGSAC 24  
 |||||:||||:||||:||||:  
 Db 74 GTCTGTGCGCGCGCGCGCGCGCAC 51

RESULT 13  
 CB605972/c  
 LOCUS  
 DEFINITION  
 AMGNMNC:NRHVS-00007-B1-A W Rat hypothalamus (10471) Rattus norvegicus cDNA clone nrhy5-00007-b11 5', mRNA sequence.  
 CB605972  
 CB605972.1 GI:29545584  
 EST.  
 Rattus norvegicus (Norway rat)  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 1 (bases 1 to 526)  
 Amgen EST Program.  
 Amgen Rat EST Program  
 Unpublished (2003)  
 Contact: Dan Fitzpatrick  
 Amgen, Inc  
 One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA  
 Tel: 805 447-4881  
 Plate: 00007 row: b column: 11.  
 Location/Qualifiers  
 1..526  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10116"  
 /clone="nrhy5-00007-b11"  
 /clone\_1lb="W Rat hypothalamus (10471)"  
 /note="Vector: pSPOR1; Site 1: SalI; Site 2: NotI; W Rat hypothalamus adult female Wistar rat avg. insert size 2.3 kb fraction 6 and 7"

ORIGIN  
 Query Match 78.3%; Score 18.8; DB 6; Length 526;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Oy 1 GTCTGTGCGCGCGTACGCGSAC 24  
 |||||:||||:||||:||||:  
 Db 1 GTCTGTGCGCGCGTACGCGSAC 24

Db 63 GTCTGTGCGCGCGCGCGCGCGCAC 40

RESULT 14  
 CK928924/c  
 LOCUS  
 DEFINITION  
 p5mgc\_000121 Normalized Magnaporthe grisea cDNA pGEM-T Easy library Magnaporthe grisea cDNA clone p5mgc\_000121, mRNA sequence.  
 CK928924  
 CK928924.1 GI:45416775  
 EST.  
 Magnaporthe grisea (anamorph: Pyricularia grisea)  
 Magnaporthe grisea  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.  
 1 (bases 1 to 538)  
 Chen, B., Li, Y., Peng, Y., Dong, H. and Li, D.  
 Large-scale identification of ESTs from Magnaporthe grisea by normalized cDNA library sequencing  
 Unpublished (2004)  
 Contact: Baoshan Chen, Youzhi Li  
 Laboratory of Subtropical Bioresource Conservation and Utilization  
 Guangxi University, China Agricultural University, Zhejiang University  
 Daxue Road 100#, Nanning, Guangxi, 530004, China  
 Tel: 0086-771-3239566  
 Fax: 0086-771-3237873  
 Email: chenbs@n.gx.cninfo.net, URL: http://www.estarray.org  
 Seq primer: M13 forward primer.  
 Location/Qualifiers  
 1..538  
 /organism="Magnaporthe grisea"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:148305"  
 /clone="p5mgc\_000121"  
 /rname\_type="Myceium", conidium, germinating conidium, swelling appressorium, mature appressorium, penetration peg  
 /dev stage="Myceium, conidium, germinating conidium, swelling appressorium, mature appressorium, penetration peg"  
 /clone\_1lb="Normalized Magnaporthe grisea cDNA pGEM-T Easy library"  
 /note="Vector: pGEM-T Easy"

ORIGIN  
 Query Match 78.3%; Score 18.8; DB 7; Length 538;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Oy 1 GTCTGTGCGCGCGTACGCGSAC 24  
 |||||:||||:||||:||||:  
 Db 177 GTCTGTGCGCGCGCGTACGCGSAC 154

RESULT 15  
 BF462782/c  
 LOCUS  
 DEFINITION  
 UI-M-CG0p-bnh-e-07-0-UI-a1 NIH BMAP Ref.4.s2 Mus musculus cDNA clone UI-M-CG0p-bnh-e-07-0-UI 3', mRNA sequence.  
 BF462782  
 BF462782.1 GI:11531965  
 EST.  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 551)  
 Bonaldo, M.F., Lennon, G. and Soares, M.B.  
 Normalization and subtraction: two approaches to facilitate gene discovery  
 Genome Res. 6 (9), 791-806 (1996)  
 97044477

PUBMED  
8889548  
Contract: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mestr@mail.nih.gov  
Oligo-OT track not found. Not a site shown in beginning of sequence  
is likely internal to the message. cDNA Library Preparation: M.B.  
Soares Lab Clone distribution: Researchers may obtain BMAP cDNA  
clones from RESEARCH GENETICS. It should be noted that Bento Soares  
is generating a small number of additional specialized  
non-redundant arrays of BMAP cDNAs whose availability will be  
considered under appropriate and limited collaborative arrangements  
Seq primer: M13 Forward  
POLYA=No.

FEATURES  
source  
1. 551  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UI-M-CGP-bnh-e-07-0-UI"  
/lab\_host="PH10B (Life Technologies)"  
/clone\_lib="NIH\_BMAP Ret4 S2"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; The  
NIH BMAP Ret4 S2 library is a subcloned library,  
ultimately derived from mouse retina tissue libraries at  
various stages of development. For a detailed description  
of the library from which this clone was derived, please  
visit our web site at brainest.eng.uiowa.edu. The tissue  
for this library was contributed by Dr. Xin-Yuan Fu, Yale  
University School of Medicine  
TAG\_SEQ=None found"

ORIGIN  
Query Match 78.3%; Score 18.8; DB 2; Length 551;  
Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCCTGTTGCGCGTACGCGSAC 24  
|||||:||||:||||:||||:  
Db 113 GTCTGCGGCGCGGACGCGCAC 90

RESULT 16  
AZ417416 577 bp DNA linear GSS 03-OCT-2000  
LOCUS IM0193N06F Mouse 10kb plasmid UGCGM library Mus musculus genomic  
DEFINITION clone UGCGM0193N06 F, genomic survey sequence.  
ACCESSION AZ417416  
VERSION AZ417416.1 GI:10541429  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 577)  
REFERENCE  
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,  
Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von  
Niederhausern, A. and Wright, D., Weiss, R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
Unpublished (2000)  
JOURNAL  
COMMENT Contact: Robert B. Weiss  
University of Utah  
Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606

Fax: 801 585 7177  
Email: ddum@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0193 Row: N Column: 06  
Seq primer: CGTGTAAACGACGCCAGT  
Class: plasmid ends  
High quality sequence stop: 577.

FEATURES  
source  
1. 577  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UGCGM0193N06"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb plasmid UGCGM library"  
/note="Vector: PWD42nv; Purified genomic DNA from M.  
musculus C57BL/6J (male) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PWD42 (gi14732114|gb|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
adapted vector DNA, and transformed into  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

ORIGIN  
Query Match 78.3%; Score 18.8; DB 8; Length 577;  
Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCCTGTTGCGCGTACGCGSAC 24  
|||||:||||:||||:||||:  
Db 411 GTCTGCGGCGCGGACGCGCAC 434

RESULT 17  
BB642652/c 610 bp mRNA linear EST 26-OCT-2001  
LOCUS BB642652 RIKEN full-length enriched, adult retina Mus musculus cDNA  
DEFINITION clone A930028E09 5', mRNA sequence.  
ACCESSION BB642652  
VERSION BB642652.1 GI:16477487  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 610)  
REFERENCE  
AUTHORS Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T.,  
Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, O.,  
Kono, H., Kondo, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,  
Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,  
Takeda, Y., Tanaka, T., Toyota, T., Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)  
Unpublished (2001)  
JOURNAL  
COMMENT Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)



prepare mouse tissues.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

#### FEATURES

source

Location/Qualifiers

1. 619

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="241004013"

/cell\_type="ES cells"

/lab\_host="GDR"

/clone\_11b="RIKEN full-length enriched, ES cells"

/note="Site 1: XhoI; Site 2: BamHI; cDNA library was

prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in

RIKEN, Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was

primed with a primer 15'

GAGAGAGAGATCTCGAGTATTAATTAATCCGCCGCCGCC 3'}. cDNA was

prepared by using trehalose thermo-activated reverse

transcriptase and subsequently enriched for full-length by

cap-trapper. cDNA went through one round of normalization

to Rot = 5.0 and subtraction to Rot = 25.0. Second strand

cDNA was prepared with the primer adapter of sequence 15'

GAGAGAGATCTCGAGTATTAATTAATCCGCCGCCGCC 3'}. "

#### ORIGIN

Query Match 78.3%; Score 18.8; DB 6; Length 619;  
Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCTGTGGCCGTCGACGCGSAC 24  
|||||:||||:||||:||||:||||:  
98 GTCTGTGGCCGTCGACGCGCAC 75

Db

#### RESULT 19

BU701308/c

UT-M-EXO-bwy-k-12-0-UI.r1 NIH\_BMAP\_EXO Mus musculus cDNA clone  
IMAGE:5704835 5', mRNA sequence.

#### ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

Tissue Procurement: Dr. James Lin, University of Iowa

cDNA library preparation: Dr. M. Bento Soares, University of Iowa

cDNA library arrayed by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/BLML at:

<http://image.llnl.gov>

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: pYX-5.

Location/Qualifiers

1. .641

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clone="IMAGR:5704835"

#### ORIGIN

Query Match 78.3%; Score 18.8; DB 5; Length 641;  
Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCTGTGGCCGTCGACGCGSAC 24  
|||||:||||:||||:||||:||||:  
119 GTCTGTGGCCGTCGACGCGCAC 96

Db

#### RESULT 20

CN721004/c

E0814B04-5 NIA Mouse four-cell-Embryo cDNA library (long) Mus  
musculus cDNA clone NIA:E0814B04 IMAGE:30906927 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1999)

Contact: Dawood B. Dudekula

Laboratory of Genetics

National Institute on Aging/National Institutes of Health

333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA

Email: [cdna@igsun.grc.nia.nih.gov](mailto:cdna@igsun.grc.nia.nih.gov)

Plate: E0814 row: B column: 04

Seq primer: M13 Reverse

High quality sequence stop: 672

POLYA=No.

#### FEATURES

source

Location/Qualifiers

1. .672

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="NIA:E0814B04 IMAGE:30906927"

/tissue\_type="4-cell stage embryo"

/dev\_stage="4-cell"

/lab\_host="DHI0B"

/clone\_lib="NIA Mouse four-cell-Embryo cDNA library (long)"  
 /note="Vector: pCMV-SPORT6 (Invitrogen); Site 1: SalI; Site 2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsn.gsc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). The mRNAs were extracted from a pool of 360 embryos at 4-cell stage. Double-stranded cDNAs were synthesized with an oligo(dT) primer [Invitrogen: 5'-pGACTGATCTAGATCGGACGGCCCTTTTCTTTT-3'] from 10.8ng of mRNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker lB-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer SalI-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2Kb. The library was constructed by Yulan Piao."

## ORIGIN

Query Match 78.3%; Score 18.8; DB 7; Length 672;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 GTCCGTGCGCCGCTACGCGSAC 24  
 |||||:||||:||||:||||:||||:  
 Db 86 GTCCGTGCGCGCCGCGGACGCGCAC 63

RESULT 21 717 bp mRNA linear EST 20-SEP-2001  
 B1732637/c LOCUS  
 DEFINITION 60355381F1 NIH\_MGC\_94 Mus musculus cDNA clone IMAGE:5362485 5',  
 mRNA sequence.  
 ACCESSION B1732637  
 VERSION B1732637.1 GI:15709650  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 717)  
 NIH-MGC http://mgs.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@bte-remail.nih.gov  
 Tissue Procurement: The Cepko Laboratory  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ILNI at:  
 http://image.llnl.gov  
 Plate: L1AM1922 row: b column: 22  
 High quality sequence stop: 646.  
 Location/Qualifiers

## FEATURES

1. 717  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5362485"  
 /tissue\_type="retina"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_94"  
 /note="Organ: eye; Vector: pCMV-SPORT6; Site\_1: NotI;

Site\_2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 3.3 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 78.3%; Score 18.8; DB 4; Length 717;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 GTCCGTGCGCCGCTACGCGSAC 24  
 |||||:||||:||||:||||:||||:  
 Db 79 GTCCGTGCGCGCCGCGGACGCGCAC 56

RESULT 22 760 bp DNA linear GSS 17-DEC-2002  
 B2550710 LOCUS  
 DEFINITION pacsl-60.2788.y2 pacsl-60 Pseudomonas aeruginosa genomic clone  
 pacsl-60.2788, genomic survey sequence.  
 ACCESSION B2550710  
 VERSION B2550710.1 GI:27154291  
 KEYWORDS GSS.  
 SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 760)  
 Spencer,J.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,  
 Burns,J.L., Kaul,R. and Olsen,M.V.  
 Whole-Genome-Sequence variation among multiple isolates of  
 Pseudomonas aeruginosa library  
 J. Bacteriol. (2002) In press  
 CONTACT: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel: 2062216954  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: Shotgun.

## FEATURES

1. 760  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="1-60"  
 /db\_xref="taxon:287"  
 /clone="pacsl-60.2788"  
 /clone\_lib="pacsl-60"  
 /note="clinical isolate 1-60 whole genomic shotgun  
 library."

## ORIGIN

Query Match 78.3%; Score 18.8; DB 8; Length 760;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 GTCCGTGCGCCGCTACGCGSAC 24  
 |||||:||||:||||:||||:||||:  
 Db 450 GTCCGTGCGCGCGGACGCGCAC 473

RESULT 23 780 bp mRNA linear EST 12-JUL-2000  
 AU080793/c LOCUS  
 DEFINITION AU080793 Sugano mouse brain mchb Mus musculus cDNA clone MIMC-6078  
 5', mRNA sequence.  
 ACCESSION AU080793  
 VERSION AU080793.1 GI:6085547  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
AUTHORS        1 (bases 1 to 780)
TITLE          Hashimoto, K., Kusuda, J., Tanuma, R., Ito, A., Hirata, M., Toyoda, A.,
              Suzuki, Y., Sasaki, W., and Sugano, S.
JOURNAL        Isolation of full-length cDNA clones from a mouse brain cDNA
COMMENT         library made by oligo-capping method
              unpublished (1999)
CONTACT        Contact: Kazuyuki Hashimoto
                Division of Genetic Resources
                National Institute of Infectious Diseases
                23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
                Email: khashi@nih.go.jp
                URL: http://www.nih.go.jp/yoken/genbank/.

FEATURES       location/Qualifiers
SOURCE         1..780
               /organism="Mus musculus"
               /mol_type="mRNA"
               /strain="C57BL"
               /db_xref="taxon:10090"
               /clone="MNCB-6078"
               /sex="female"
               /dev_stage="adult"
               /lab_host="TOP10"
               /clone_lib="Sugano mouse brain mncb"
               /note="Organ: brain; Vector: pME18S-FL3; 1st strand cDNA
               was primed with an oligo(dT) primer
               ATGGGCGCTTTTATTTTTTTT"; double-stranded cDNA was
               ligated to a DraIII adaptor [TGTTGGCTACTAG], digested and
               cloned into distinct DraIII sites of the pME18S-FL3. XhoI
               sites just outside the DraIII sites can be used to isolate
               the cDNA insert. Size selection was performed to exclude
               fragments <1.5 kb. Library was constructed by Sugano et
               al. (University of Tokyo, Institute of Medical Science).
               Custom primer used for sequencing: 5' end primer
               [CTCTGTGCTTAAGAAGTCGC], 3' end primer
               [CGACGTGCACTGCAGCACAA]"

ORIGIN
Query Match      78.3%; Score 18.8; DB 1; Length 780;
Best Local Similarity 70.8%; Pred. No. 2.7e+03;
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY
1 GTCCGTGTCGCCGCTSACSCGSAC 24
|||||:|||||:~::~|||:|
Db
94 GTCCGTGGGGCGCGGGGACGCGCAC 71

RESULT 24
BZ565923/contig1 782 bp DNA linear GSS 17-DEC-2007
LOCUS BZ565923/contig1 782 bp DNA linear GSS 17-DEC-2007
DEFINITION pac62-164_5820.y2 pac62-164 Pseudomonas aeruginosa genomic clone
ACCESSION BZ565923
VERSION pac62-164_5820, genomic survey sequence.
KEYWORDS BZ565923.1 GI:27194532
SOURCE GSS.
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 782)
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
Whole-genome-sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol. (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

```

	FEATURES	Location/Qualifiers 1..782
	SOURCE	/organism="Pseudomonas aeruginosa" /mol_type="genomic DNA" /strain="C57BL"
		/db_xref="taxon:287"
		/clone_1lib="pac62-164_S820"
		/notes="clinical isolate 2-164 Whole genomic shotgun library."
	ORIGIN	
	Query Match	78.3%; Score 18.8; DB 8; Length 782;
	Best Local Similarity	70.8%; Pred. No. 2.7e+03;
	Matches	17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
OY	1 GTCCCTGTGCGCGTACSGCSAC 24	
Dn	::   ::   ::   ::	
	Db	551 GTCCGCCTGGCGCGTCACCACC 528
RESULT 25	AU080514	793 bp mRNA linear EST 12-JUN-2000
Locus	AU080514/c	
DEFINITION	AU080514 Sugano mouse brain mncb Mus musculus CDNA clone MNCD-5748 5', mRNA sequence.	
ACCESSION	AU080514	
VERSION	AU080514.1 GI:6085268	
KEYWORDS	EST.	
SOURCE	Mus musculus (house mouse)	
ORGANISM	Mus musculus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. I (Bases 1 to 793)	
REFERENCE	Hashimoto,K., Kusuda,U., Tanuma,R., Ito,A., Hirata,M., Toyoda,A., Suzuki,Y., Sasaki,M. and Sugano,S. Isolation of full-length cDNA clones from a mouse brain cDNA library made by oligo-capping method unpublished (1999) Contact: Katsuyuki Hashimoto Division of Genetic Resources National Institute of Infectious Diseases 23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan Email: khashim@nih.go.jp URL: http://www.nih.go.jp/yoken/genbank/. Location/Qualifiers 1..793	
JOURNAL COMMENT	/organism="Mus musculus" /mol_type="mRNA" /_strain="C57BL" /_db_xref="taxon:10090" /_clone="MNCD-5748" /_sex="female" /_dev_stage="adult" /_lab_host="TOP10" /_clone_lib="Sugano mouse brain mncb" /_note="Organ: Brain; Vector: pHE18S-FL3; 1st strand cDNA was primed with an oligo (dT) primer ATGCCGCCCCTTTTTTT"; double-stranded cDNA was digested another [ATGTCGACTAGTG], cloned into distinct DraIII sites of the pHE18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5 kb. Library was constructed by Sugano et al.(University of Tokyo, Institute of Medical Science). Cesium primer used for sequencing: 5' end primer [CTTCTGCTCAAAAGCTGG] , 3' end primer [CGACTGCAGCTGAACA]	
FEATURES	SOURCE	
ORIGIN		
Query Match	78.3%; Score 18.8; DB 1; Length 793;	
Best Local Similarity	70.8%; Pred. No. 2.7e+03;	
Matches	17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;	

Qy 1 GTCTGTGGCGCGTGTACGCGSAC 24  
 Db 95 GTCTGTGGCGCGCGGAGACGCGCAC 72

RESULT 26  
 CNS077KA/c 907 bp DNA linear GSS 08-JUL-2001  
 LOCUS T7 end of clone BB0A012E03 of library BB0A from strain CBS 4732  
 DEFINITION of *Pichia angusta*, genomic survey sequence.  
 ACCESSION AL432800  
 VERSION AL432800.1 GI:12216214  
 KEYWORDS GSS.  
 ORGANISM *Pichia angusta*  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.

REFERENCE 1 (bases 1 to 907)  
 AUTHORS Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bojotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B., Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Pottier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P., and Weissenbach, J.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies  
 JOURNAL FEMS Lett. 487 (1), 3-12 (2000)  
 MEDLINE 20584711  
 PUBMED 11152876  
 REFERENCE 2 (bases 1 to 907)  
 AUTHORS Blandin, G., Llorente, B., Malpertuy, A., Wincker, P., Artiguenave, F., and Dujon, B.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 13. *Pichia angusta*  
 JOURNAL FEMS Lett. 487 (1), 76-81 (2000)  
 MEDLINE 20584723  
 PUBMED 11152888  
 REFERENCE 3 (bases 1 to 907)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage, 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: seqref@genoscope.cns.fr - Web: www.genoscope.cns.fr)  
 COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces exiguus*, *Saccharomyces servazii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES  
 source  
 1..907  
 /organism="Pichia angusta"  
 /mol\_type="genomic DNA"  
 /strain="CBS 4732"  
 /db\_xref="taxon:4905"  
 /clone="BB0A012E03"  
 /clone\_1lb="BB0AA"  
 /note="end : 72"

ORIGIN  
 Query Match 78.3%; Score 18.8; DB 9; Length 907;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCTGTGGCGCGTGTACGCGSAC 24  
 Db 302 GTCTGTGGCGCGGAGACGCGCAC 279

RESULT 27  
 B2568254/c 909 bp DNA linear GSS 17-DEC-2002  
 LOCUS paces2-164\_7441.y2 paces2-164 *Pseudomonas aeruginosa* genomic clone  
 DEFINITION paces2-164\_7441, genomic survey sequence.  
 ACCESSION B2568254  
 VERSION B2568254.1 GI:27200910  
 KEYWORDS GSS.  
 SOURCE *Pseudomonas aeruginosa*  
 ORGANISM *Pseudomonas aeruginosa*  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 909)  
 AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.B., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R., and Olsen, M.V.  
 TITLE Whole-Genome-Sequence variation among multiple isolates of *Pseudomonas aeruginosa* library  
 J. Bacteriol. (2002) in press  
 COMMENT Contact: Chris K. Raymond  
 Genome Center  
 University of Washington  
 Box 352145, Seattle, WA 98105-2145, USA  
 Tel.: 2062216854  
 Fax: 2066857244  
 Email: craymond@u.washington.edu  
 Class: shotgun.

FEATURES  
 source  
 1..909  
 /organism="Pseudomonas aeruginosa"  
 /mol\_type="genomic DNA"  
 /strain="2-164"  
 /db\_xref="taxon:287"  
 /clone="paces2-164\_7441"  
 /clone\_1lb="paces2-164"  
 /note="clinical isolate 2-164 whole genomic shotgun library."

ORIGIN  
 Query Match 78.3%; Score 18.8; DB 8; Length 909;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCTGTGGCGCGTGTACGCGSAC 24  
 Db 348 GTCTGTGGCGCGCACCGCGCAC 325

RESULT 28  
 CNS079PU/c 927 bp DNA linear GSS 08-JUL-2001  
 LOCUS T3 end of clone BB0A031C05 of library BB0A from strain CBS 4732  
 DEFINITION of *Pichia angusta*, genomic survey sequence.  
 ACCESSION AL435592  
 VERSION AL435592.1 GI:12219005  
 KEYWORDS GSS.  
 SOURCE *Pichia angusta*  
 ORGANISM *Pichia angusta*  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.

REFERENCE 1 (bases 1 to 927)  
 AUTHORS Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bojotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B., Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Pottier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P., and Weissenbach, J.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies  
 JOURNAL FEMS Lett. 487 (1), 3-12 (2000)  
 MEDLINE 20584711  
 PUBMED 11152876  
 REFERENCE 2 (bases 1 to 927)



**AUTHORS** Blandin, G., Llorente, B., Malpertuy, A., Wincker, P., Artiguenave, F. and Dujon, B.  
**TITLE** Genomic exploration of the hemiascomycetous yeasts: 13. *Pichia angusta*  
**JOURNAL** FEBS Lett. 487 (1), 76-81 (2000)  
**MEDLINE** 20584723  
**PUBMED** 11152868  
**REFERENCE** 3 (bases 1 to 927)  
**AUTHORS** Genoscope.  
**TITLE** Direct Submision  
**JOURNAL** Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : beqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
**COMMENT** This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvatum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.  
**FEATURES** location/Qualifiers  
 source  
 1..927  
 /organism="Pichia angusta"  
 /mol\_type="genomic DNA"  
 /strain="CBS 4732"  
 /db\_xref="taxon:4905"  
 /clone="BB0AA031C05"  
 /clone\_1lb="BB0AA"  
 /note="end : T3"  
**ORIGIN**  
 Query Match 78.3%; Score 18.8; DB 9; Length 927;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 1 GTCTGTGTCGCGCGTACGCGSAC 24  
 |||||:||||:||||:||||:  
 190 GTCTGTGTCGCGCGAGACCGCGAC 167  
**RESULT 29** 1024 bp DNA linear GSS 08-JUL-2001  
**LOCUS** CNS0795W  
**DEFINITION** T3 end of clone BB0AA026E07 of library BB0AA from strain CBS 4732 of *Pichia angusta*, genomic survey sequence.  
**ACCESSION** AL434874  
**VERSION** AL434874.1 GI:12218287  
**KEYWORDS** GSS.  
**SOURCE** *Pichia angusta*  
**ORGANISM** *Pichia angusta*  
**REFERENCE** Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; *Pichia*.  
**AUTHORS** 1 (bases 1 to 1024)  
 Soucier, J. L., Aigle, M., Artiguenave, F., Blandin, G., Bolicin-Fukuhara, M., Bon, E., Broctier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Leplingle, A., Llorente, B., Malpertuy, A., Neveglise, C., Olier-Kalogeropoulos, O., Potier, S., Sautin, W., Tekala, F., Toffano-Nicche, C., Wesolowski-Louvel, M., Wincker, P. and Weissensbach, J.  
**TITLE** Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies  
**JOURNAL** FEBS Lett. 487 (1), 3-12 (2000)  
**MEDLINE** 20584711  
**PUBMED** 11152876  
**REFERENCE** 2 (bases 1 to 1024)  
**AUTHORS** Blandin, G., Llorente, B., Malpertuy, A., Wincker, P., Artiguenave, F. and Dujon, B.  
**TITLE** Genomic exploration of the hemiascomycetous yeasts: 13. *Pichia angusta*  
**JOURNAL** FEBS Lett. 487 (1), 76-81 (2000)

**MEDLINE** 20584723  
**PUBMED** 11152868  
**REFERENCE** 3 (bases 1 to 1024)  
**AUTHORS** Genoscope.  
**TITLE** Direct Submision  
**JOURNAL** Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : beqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
**COMMENT** This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvatum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.  
**FEATURES** location/Qualifiers  
 source  
 1..1024  
 /organism="Pichia angusta"  
 /mol\_type="genomic DNA"  
 /strain="CBS 4732"  
 /db\_xref="taxon:4905"  
 /clone="BB0AA026E07"  
 /clone\_1lb="BB0AA"  
 /note="end : T3"  
**ORIGIN**  
 Query Match 78.3%; Score 18.8; DB 9; Length 1024;  
 Best Local Similarity 70.8%; Pred. No. 2.7e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 1 GTCTGTGTCGCGCGTACGCGSAC 24  
 |||||:||||:||||:||||:  
 65 GTCTGTGTCGCGCGAGACCGCGAC 88  
**RESULT 30** 1325 bp mRNA linear HTC 03-APR-2004  
**LOCUS** AK010405/c  
**DEFINITION** Mus musculus ES cells cDNA, RIKEN full-length enriched library, clone:2410004013 product:porcupine homolog (Drosophila), full insert sequence.  
**ACCESSION** AK010405  
**VERSION** AK010405.1 GI:12845824  
**KEYWORDS** HTC; CAP trapper.  
**SOURCE** Mus musculus (house mouse)  
**ORGANISM** Mus musculus  
**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
**AUTHORS** 1  
 Carninci, P. and Hayashizaki, Y.  
**TITLE** High-efficiency full-length cDNA cloning  
**JOURNAL** Meth. Enzymol. 303, 19-44 (1999)  
**MEDLINE** 99279253  
**PUBMED** 10349636  
**REFERENCE** 2  
 Carninci, P. and Hayashizaki, Y.  
**AUTHORS** Kono, H., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
**TITLE** Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
**JOURNAL** Genome Res. 10 (10), 1617-1630 (2000)  
**MEDLINE** 20499374  
**PUBMED** 11042159  
**REFERENCE** 3  
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, U., Nishikawa, K., Kikunishi, T., Yasuiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, B., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.



TITLE	JOURNAL MEDLINS PUBLISHED REFERENCE AUTHORS	TITLE	JOURNAL MEDLINS PUBLISHED REFERENCE AUTHORS
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexillary sequencer Genome Res. 10 (11), 1757-1771 (2000)	20530913 11076861	The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001)	5
The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) (bases 1 to 1325)	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arkawa,T., Bono,H., Carninci,P., Fukuda,S., Fujinishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiwaku,T., Horii,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kikukawa,T., Kato,H., Kawaji,Y., Kohjima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Saeki,D., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toyoi,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.	Direct Submission Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gscc.riken.jp, URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216]	Please visit our web site ( <a href="http://genome.gsc.riken.jp/">http://genome.gsc.riken.jp/</a> ) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGACAGATCCAGAGCTCTTTTCTTTTTTTTAA 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGACTTCGAGTAATTAAATTATATCCCCCCCCC 3']. cDNA was cleaved with XhoI and SetI. Cloning sites, 5' end: XhoI; 3' end: SetI. Host: SOLR.	Location/Qualifiers 1..1325 /organism="Mus musculus" /mol_type="mRNA" /strain="CS7BL/6J" /db_xref="PANTOM DB:2410004O13" /db_xref="taxon:10090" /clone="2410004O13" /cell_type="ES cells" /clone_1ib="RIKEN full-length enriched mouse cDNA library" 315..998 /note="unnamed protein product; porcupine homolog (Drosophila) (MGP MGI:1890212) putative"	/codon_start=1 /protein_id="BAB26914.1" /db_xref="gi:12845825" /translat="MAQDAARVAVYESAVSFHNSYFVGELSEATATLAGAPTEKRD HLNDIVTSRLNVELPSRWVEVTSNMLPWSYVNMYFNALRGTFSSAVLYTAA SALHGSPFLAVALISGLAFITYEHVLRKLADILSACLISKECLDDCGRRHIGIG VRANLPFGALATPHLSLGSLFDVDVDTTBEGGYGMATYVKMSLSMASHHYTGC CWITPRILIG"	

polyA_signal	1303..1308	/note="putative"
polyA_site	1325	/note="putative"
ORIGIN		
Query Match	78.3%;	Score 18.8; DB 3;
Best Local Similarity	70.8%;	Pred. No. 2.6e+03;
Matches 17; Conservative	5; Mismatches	2; Indels 0; Gaps 0;
Oy	1 GTCCCTGATGCCCGTSACSCGSAC 24	:    :    :    :
Db	98 GTCCCTGGGGGCCCGGAGACGCCAC 75	

Search completed: July 20, 2005, 18:16:21  
Job time : 3014.43 secs

**THIS PAGE IS BLANK**



C 94	16.8	70.0	1230	6	ABQ79850	Abq79850 Human TAN	C 167	16.8	70.0	2037	8	ACA68485	ACA68485 Novel hum
C 95	16.8	70.0	1230	12	AD133308	Ad133308 Partial h	C 168	16.8	70.0	2037	8	ABX98950	ABX98950 Novel hum
C 96	16.8	70.0	1254	13	ADT42027	Adt42027 Bacterial	C 169	16.8	70.0	2037	8	ACC81427	ACC81427 Human sec
C 97	16.8	70.0	1263	3	AA251245	AA251245 Human TAN	C 170	16.8	70.0	2037	8	ACA95751	ACA95751 Novel hum
C 98	16.8	70.0	1263	6	ABQ79849	Abq79849 Human TAN	C 171	16.8	70.0	2037	8	ACD04669	ACD04669 Novel hum
C 99	16.8	70.0	1263	12	AD133307	Ad133307 Partial h	C 172	16.8	70.0	2037	8	ACC88110	ACC88110 Human sec
C 100	16.8	70.0	1263	13	ADSA5825	Adsa5825 Bacterial	C 173	16.8	70.0	2037	8	ACF12772	ACF12772 Human sec
C 101	16.8	70.0	1476	11	ABD16583	Abd16583 Pseudomon	C 174	16.8	70.0	2037	8	ACA96487	ACA96487 Human PRO
C 102	16.8	70.0	1707	8	ADA69639	Ada69639 Rice gene	C 175	16.8	70.0	2037	8	ACA65261	ACA65261 Human PRO
C 103	16.8	70.0	1797	11	ABD15840	Abd15840 Pseudomon	C 176	16.8	70.0	2037	8	ACA73987	ACA73987 Human sec
C 104	16.8	70.0	1887	11	ABD17552	Abd17552 Pseudomon	C 177	16.8	70.0	2037	8	ACA74399	ACA74399 Novel hum
C 105	16.8	70.0	2025	11	ABD16325	Abd16325 Pseudomon	C 178	16.8	70.0	2037	8	ACA96794	ACA96794 Human PRO
C 106	16.8	70.0	2037	4	AA546220	AA546220 Human DNA	C 179	16.8	70.0	2037	8	ACD10900	ACD10900 CDNA enco
C 107	16.8	70.0	2037	8	ACA89670	ACA89670 CDNA enco	C 180	16.8	70.0	2037	8	ACC91596	ACC91596 Human sec
C 108	16.8	70.0	2037	8	ACA73680	ACA73680 Human sec	C 181	16.8	70.0	2037	8	ACD02931	ACD02931 CDNA enco
C 109	16.8	70.0	2037	8	ACA05995	ACA05995 Human sec	C 182	16.8	70.0	2037	8	ACC87496	ACC87496 Human sec
C 110	16.8	70.0	2037	8	ACA66829	ACA66829 CDNA enco	C 183	16.8	70.0	2037	8	ACC86080	ACC86080 Human sec
C 111	16.8	70.0	2037	8	ACF20404	Acf20404 Human sec	C 184	16.8	70.0	2037	8	ACA65568	ACA65568 Human PRO
C 112	16.8	70.0	2037	8	ACF19790	Acf19790 Human sec	C 185	16.8	70.0	2037	8	ACA94385	ACA94385 Human PRO
C 113	16.8	70.0	2037	8	ACD22078	ACD22078 Human sec	C 186	16.8	70.0	2037	8	ACA98129	ACA98129 Human PRO
C 114	16.8	70.0	2037	8	ACF13243	ACf13243 Human sec	C 187	16.8	70.0	2037	8	ACA91631	ACA91631 Novel hum
C 115	16.8	70.0	2037	8	ACD25346	ACd25346 Human sec	C 188	16.8	70.0	2037	8	ACA90845	ACA90845 Novel hum
C 116	16.8	70.0	2037	8	ACF00395	ACf00395 Human sec	C 189	16.8	70.0	2037	8	ACD16392	ACD16392 Human sec
C 117	16.8	70.0	2037	8	ACA72452	ACA72452 Novel hum	C 190	16.8	70.0	2037	8	ACD17553	ACD17553 Human sec
C 118	16.8	70.0	2037	8	ACD04976	ACd04976 Novel hum	C 191	16.8	70.0	2037	8	ACC92210	ACC92210 Human sec
C 119	16.8	70.0	2037	8	ACD18437	ACd18437 Human sec	C 192	16.8	70.0	2037	8	ACA75067	ACA75067 CDNA enco
C 120	16.8	70.0	2037	8	ACD08444	ACd08444 Human sec	C 193	16.8	70.0	2037	8	ACA91938	ACA91938 Human PRO
C 121	16.8	70.0	2037	8	ACA88878	ACA88878 Novel hum	C 194	16.8	70.0	2037	8	ACA71582	ACA71582 Human sec
C 122	16.8	70.0	2037	8	ACA70320	ACA70320 Human sec	C 195	16.8	70.0	2037	8	ACC90982	ACC90982 Human sec
C 123	16.8	70.0	2037	8	ACD12542	ACd12542 Novel hum	C 196	16.8	70.0	2037	8	ACA65992	ACA65992 CDNA enco
C 124	16.8	70.0	2037	8	ACC74457	ACC74457 Human sec	C 197	16.8	70.0	2037	8	ACA95137	ACA95137 CDNA enco
C 125	16.8	70.0	2037	8	ACD16085	ACd16085 Human sec	C 198	16.8	70.0	2037	8	ACD16699	ACD16699 Human sec
C 126	16.8	70.0	2037	8	ACD25653	ACd25653 Novel hum	C 199	16.8	70.0	2037	8	ACD15778	ACD15778 Human sec
C 127	16.8	70.0	2037	8	ACD18130	ACd18130 Human sec	C 200	16.8	70.0	2037	8	ABX16881	ABX16881 Human CDN
C 128	16.8	70.0	2037	8	ACC88417	ACC88417 Human sec	C 201	16.8	70.0	2037	9	ACA97822	ACA97822 Novel hum
C 129	16.8	70.0	2037	8	ACD11771	ACd11771 Human sec	C 202	16.8	70.0	2037	9	ACA99271	ACA99271 Novel hum
C 130	16.8	70.0	2037	8	ACD18838	ACd18838 Human sec	C 203	16.8	70.0	2037	9	ACC91903	ACC91903 Human sec
C 131	16.8	70.0	2037	8	ABX98448	ABx98448 Human CDN	C 204	16.8	70.0	2037	9	ACD11314	ACD11314 Novel hum
C 132	16.8	70.0	2037	8	ACD14199	ACd14199 Human PRO	C 205	16.8	70.0	2037	9	ACD15164	ACD15164 Human sec
C 133	16.8	70.0	2037	8	ACD09979	ACd09979 Human sec	C 206	16.8	70.0	2037	9	ACD11928	ACD11928 Human sec
C 134	16.8	70.0	2037	8	ACC88724	ACC88724 Human sec	C 207	16.8	70.0	2037	9	ACC96057	ACC96057 Human sec
C 135	16.8	70.0	2037	8	ACD21464	ACd21464 Human sec	C 208	16.8	70.0	2037	9	ACF16620	ACF16620 Human sec
C 136	16.8	70.0	2037	8	ABX75836	ABx75836 Human CDN	C 209	16.8	70.0	2037	9	ACF02738	ACF02738 Human sec
C 137	16.8	70.0	2037	8	ABX98039	ABx98039 Human PRO	C 210	16.8	70.0	2037	9	ACF03045	ACF03045 Human sec
C 138	16.8	70.0	2037	8	ACA57978	ACA57978 Novel hum	C 211	16.8	70.0	2037	9	ACF21632	ACF21632 Human sec
C 139	16.8	70.0	2037	8	ACA57978	ACA57978 Novel hum	C 212	16.8	70.0	2037	9	ACF10316	ACF10316 Human sec
C 140	16.8	70.0	2037	8	ACD14506	ACd14506 Human PRO	C 213	16.8	70.0	2037	9	ACF78209	ACF78209 Human sec
C 141	16.8	70.0	2037	8	ACC91289	ACC91289 Human sec	C 214	16.8	70.0	2037	9	ACD46914	ACD46914 Human sec
C 142	16.8	70.0	2037	8	ACC89031	ACC89031 Human sec	C 215	16.8	70.0	2037	9	ACD49677	ACD49677 Human sec
C 143	16.8	70.0	2037	8	ACD07228	ACd07228 Human PRO	C 216	16.8	70.0	2037	9	ACP28444	ACP28444 Human sec
C 144	16.8	70.0	2037	8	ACA67679	ACA67679 Human PRO	C 217	16.8	70.0	2037	9	ACD89134	ACD89134 Human sec
C 145	16.8	70.0	2037	8	ACC81734	ACC81734 Human sec	C 218	16.8	70.0	2037	9	ACD84529	ACD84529 Human PRO
C 146	16.8	70.0	2037	8	ACC89338	ACC89338 Human sec	C 219	16.8	70.0	2037	9	ACD99303	ACD99303 CDNA enco
C 147	16.8	70.0	2037	8	ACC86694	ACC86694 Human sec	C 220	16.8	70.0	2037	9	ADA78343	ADA78343 Human sec
C 148	16.8	70.0	2037	8	ACC89992	ACC89992 Human sec	C 221	16.8	70.0	2037	9	ACF49045	ACF49045 Human sec
C 149	16.8	70.0	2037	8	ACC93131	ACC93131 Human sec	C 222	16.8	70.0	2037	9	ACD09365	ACD09365 Human sec
C 150	16.8	70.0	2037	8	ACA72759	ACA72759 Human PRO	C 223	16.8	70.0	2037	9	ACF12158	ACF12158 Human sec
C 151	16.8	70.0	2037	8	ACA89277	ACA89277 Human sec	C 224	16.8	70.0	2037	9	ACF41392	ACF41392 Human sec
C 152	16.8	70.0	2037	8	ACA70013	ACA70013 Human sec	C 225	16.8	70.0	2037	9	ACF16006	ACF16006 Human sec
C 153	16.8	70.0	2037	8	ACA97156	ACA97156 Novel hum	C 226	16.8	70.0	2037	9	ACF16313	ACF16313 Human sec
C 154	16.8	70.0	2037	8	ACA91152	ACA91152 Novel hum	C 227	16.8	70.0	2037	9	ACD32140	ACD32140 Human sec
C 155	16.8	70.0	2037	8	ACA70934	ACA70934 Human sec	C 228	16.8	70.0	2037	9	ACF18948	ACF18948 Human sec
C 156	16.8	70.0	2037	8	ACA95444	ACA95444 Novel hum	C 229	16.8	70.0	2037	9	ACF09395	ACF09395 Human sec
C 157	16.8	70.0	2037	8	ACC86387	ACC86387 Human sec	C 230	16.8	70.0	2037	9	ACF78516	ACF78516 Human sec
C 158	16.8	70.0	2037	8	ACC90259	ACC90259 Human sec	C 231	16.8	70.0	2037	9	ACF52115	ACF52115 Human sec
C 159	16.8	70.0	2037	8	ACD12867	ACd12867 Human sec	C 232	16.8	70.0	2037	9	ACF26602	ACF26602 Human sec
C 160	16.8	70.0	2037	8	ACP20097	ACP20097 Human sec	C 233	16.8	70.0	2037	9	ACF24395	ACF24395 Human sec
C 161	16.8	70.0	2037	8	ABX77041	ABx77041 Human PRO	C 234	16.8	70.0	2037	9	ACF63706	ACF63706 Human sec
C 162	16.8	70.0	2037	8	ACA73373	ACA73373 Novel hum	C 235	16.8	70.0	2037	9	ACF50580	ACF50580 Human sec
C 163	16.8	70.0	2037	8	ACA68916	ACA68916 Novel hum	C 236	16.8	70.0	2037	9	ACH08051	ACH08051 Human sec
C 164	16.8	70.0	2037	8	ACA74760	ACA74760 CDNA enco	C 237	16.8	70.0	2037	9	ACF13857	ACF13857 Human sec
C 165	16.8	70.0	2037	8	ACA70627	ACA70627 Human sec	C 238	16.8	70.0	2037	9	ACD41783	ACD41783 Human sec
C 166	16.8	70.0	2037	8	ACD14813	ACd14813 Human PRO	C 239	16.8	70.0	2037	9	ACP32196	ACP32196 Human sec

C 240	16.8	70.0	2037	9	ACF23474	Act23474 Human sec
C 241	16.8	70.0	2037	9	ACF40164	Act40164 Human sec
C 242	16.8	70.0	2037	9	ACD45686	Act45686 Human sec
C 243	16.8	70.0	2037	9	ACF53343	Act53343 Human sec
C 244	16.8	70.0	2037	9	ACF25723	Act25723 Human sec
C 245	16.8	70.0	2037	9	ACF45361	Act45361 Human sec
C 246	16.8	70.0	2037	9	ACF29799	Act29799 Human sec
C 247	16.8	70.0	2037	9	ACD90055	Act90055 Human sec
C 248	16.8	70.0	2037	9	ACD84836	Act84836 Human PRO
C 249	16.8	70.0	2037	9	ACD98996	Act98996 CDNA enco
C 250	16.8	70.0	2037	9	ACF77288	Act77288 Human sec
C 251	16.8	70.0	2037	9	ACF76981	Act76981 Human sec
C 252	16.8	70.0	2037	9	ACF49966	Act49966 Human sec
C 253	16.8	70.0	2037	9	ACF50273	Act50273 Human sec
C 254	16.8	70.0	2037	9	ACD09672	Act09672 Human sec
C 255	16.8	70.0	2037	9	ACD08751	Act08751 Human sec
C 256	16.8	70.0	2037	9	ACF12465	Act12465 Human sec
C 257	16.8	70.0	2037	9	ACG94973	Act94973 Human sec
C 258	16.8	70.0	2037	9	ACD2692	Act2692 Human sec
C 259	16.8	70.0	2037	9	ACF15392	Act15392 Human sec
C 260	16.8	70.0	2037	9	ACG97487	Act97487 Human sec
C 261	16.8	70.0	2037	9	ACG92517	Act92517 Human sec
C 262	16.8	70.0	2037	9	ACF14164	Act14164 Human sec
C 263	16.8	70.0	2037	9	ACF14471	Act14471 Human sec
C 264	16.8	70.0	2037	9	ACF09702	Act09702 Human sec
C 265	16.8	70.0	2037	9	ACD45993	Act45993 Human sec
C 266	16.8	70.0	2037	9	ACD48142	Act48142 Human sec
C 267	16.8	70.0	2037	9	ACD67873	Act67873 CDNA enco
C 268	16.8	70.0	2037	9	ACF25681	Act25681 Human sec
C 269	16.8	70.0	2037	9	ACF2565	Act2565 Human sec
C 270	16.8	70.0	2037	9	ACD85143	Act85143 Human sec
C 271	16.8	70.0	2037	9	ACD84222	Act84222 Human PRO
C 272	16.8	70.0	2037	9	ACD88213	Act88213 Human sec
C 273	16.8	70.0	2037	9	ACF09000	Act09000 Human sec
C 274	16.8	70.0	2037	9	ACF32503	Act32503 Human sec
C 275	16.8	70.0	2037	9	ACH12163	Act12163 CDNA enco
C 276	16.8	70.0	2037	9	ACH12470	Act12470 CDNA enco
C 277	16.8	70.0	2037	9	ACD40862	Act40862 Human sec
C 278	16.8	70.0	2037	9	ACF18334	Act18334 Human sec
C 279	16.8	70.0	2037	9	ACF08781	Act08781 Human sec
C 280	16.8	70.0	2037	9	ACF31582	Act31582 Human sec
C 281	16.8	70.0	2037	9	ACF52422	Act52422 Human sec
C 282	16.8	70.0	2037	9	ACD50291	Act50291 Human sec
C 283	16.8	70.0	2037	9	ACF38994	Act38994 Human sec
C 284	16.8	70.0	2037	9	ACF26909	Act26909 Human sec
C 285	16.8	70.0	2037	9	ACF25009	Act25009 Human sec
C 286	16.8	70.0	2037	9	ACF4589	Act4589 Human sec
C 287	16.8	70.0	2037	9	ACF28137	Act28137 Human sec
C 288	16.8	70.0	2037	9	ACD89441	Act89441 Human sec
C 289	16.8	70.0	2037	9	ACF64013	Act64013 Human sec
C 290	16.8	70.0	2037	9	ACF60653	Act60653 Human sec
C 291	16.8	70.0	2037	9	ACH12777	Act12777 CDNA enco
C 292	16.8	70.0	2037	9	ACH10200	Act10200 Human sec
C 293	16.8	70.0	2037	9	ACD04055	Act04055 Human sec
C 294	16.8	70.0	2037	9	ACD10593	Act10593 Human sec
C 295	16.8	70.0	2037	9	ACF42620	Act42620 Human sec
C 296	16.8	70.0	2037	9	ACF18641	Act18641 Human sec
C 297	16.8	70.0	2037	9	ACF02431	Act02431 Human sec
C 298	16.8	70.0	2037	9	ACF21939	Act21939 Human sec
C 299	16.8	70.0	2037	9	ACF10623	Act10623 Human sec
C 300	16.8	70.0	2037	9	ACF34075	Act34075 Human sec

## ALIGNMENTS

RESULT 1  
ID AEN86367 standard; DNA; 24 BP.  
XX AC AEN86367;  
XX DT 21-OCT-2002 (first entry)

XX DE S. coelicolor scdA gene internal segment amplifying primer 2.  
XX Antibiotic; bacterium; scdA; afaA; scdR; arpA; barA; actinorhodine; Act;  
KM undecylprodigiosin; Red; PCR; primer; ss.  
XX OS Streptomyces coelicolor.  
XX CA2322241-A1.  
XX PD 23-APR-2002.  
XX PF 23-OCT-2000; 2000CA-02322241.  
XX PR 23-OCT-2000; 2000CA-02322241.  
XX (PLAN-) PLANT BIOSCIENCE LTD.  
XX PA Takano E, Bibb M;  
XX WPI; 2002-501089/54.  
XX DR  
XX PT Modifying antibiotic-producing Streptomyces, to increase, or alter timing  
XX of, antibiotic production, by deleting the scdA or scdR genes.  
XX PS Claim 19; Page 53; 64pp; English.  
XX SS  
XX CC The invention provides a method for modifying an antibiotic-producing  
CC strain of Streptomyces to increase production of antibiotics or to alter  
CC the timing of antibiotic production. The modification is functional  
CC deletion of the scdA gene of S. coelicolor, or its homologues, but is not  
CC functional deletion of the afaA gene of S. griseus, or the modification is  
CC functional deletion of the scdR gene of S. coelicolor, or its homologues,  
CC but is not deletion of any of S. griseus nor any of S. viridnatae. The  
CC method is particularly used for production of the antibiotics  
CC actinorhodine (Act) and undecylprodigiosin (Red). The present sequence  
CC represents a PCR primer for amplifying an internal segment of the scdA  
CC gene from S. coelicolor M145 total DNA  
XX  
SQ Sequence 24 BP; 2 A; 7 C; 6 G; 4 T; 0 U; 5 Other;  
Query Match 91.7%; Score 22; DB 6; Length 24;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCTGCTGCGCGCTGACGCGAC 24  
Db 1 GTCTGCTGCGCGCTGACGCGAC 24  
RESULT 2  
ID ADE14796 standard; DNA; 24 BP.  
XX AC ADE14796;  
XX DT 29-JAN-2004 (first entry)  
XX DE PCR primer 2 related to Streptomyces coelicolor antibiotic production.  
XX KM antibiotic-producing strain; antibiotic production; scdA gene; scdR gene;  
KM actinorhodin; undecylprodigiosin; PCR; primer; ss.  
XX OS Synthetic.  
XX OS Streptomyces coelicolor.  
XX US2003124644-A1.  
XX PD 03-JUL-2003.  
XX PF 23-OCT-2001; 2001US-00017471.  
XX PR 23-OCT-2000; 2000US-0242561P.

```
XX (TAKA/) TAKANO E.  
PA (BIBB/) BIBB M J.  
XX  
XX Takano E, Bibb MJ;  
XX WPI, 2003-810983/76.  
XX  
XX Modifying an antibiotic-producing strain of Streptomyces coelicolor or  
XX Streptomyces lividans to increase or alter the timing of antibiotic  
XX production in the strain, comprises functionally deleting in the strain  
XX the scbA or ScbR gene.  
XX  
XX Example 8; Page 9; 33pp; English.  
XX  
XX This invention relates to the novel modification of an antibiotic-  
XX producing strain of Streptomyces coelicolor or Streptomyces lividans to  
XX increase or to alter the timing of antibiotic production in the strain  
XX CC The method comprises functionally deleting in the strain the scbA or ScbR  
XX CC gene. The method is useful in increasing and altering the timing of  
XX CC antibiotic production (especially actinorhodin and undecylprodigiosin) in  
XX CC Streptomyces species, particularly Streptomyces coelicolor or  
XX CC Streptomyces lividans. The present sequence is that of a degenerate PCR  
XX CC primer which was used for amplification of a region of the Streptomyces  
XX CC coelicolor ScbA gene during the exemplification of the invention.  
XX  
XX Sequence 24 BP; 2 A; 7 C; 6 G; 4 T; 0 U; 5 Other;  
XX  
XX Query Match 91.7%; Score 22; DB 10; Length 24;  
XX Best Local Similarity 79.2%; Pred. No. 32;  
XX Matches 19; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX  
XX 1 GTCTGTGCGCGTSGACGCGSAC 24  
XX 1 GTCTGTGCGCGCGTSGACGCGSAC 24  
XX  
XX Db  
XX  
XX RESULT 3  
XX ABBN6379  
XX ID ABBN6379 standard; DNA; 4346 BP.  
XX  
XX ABBN6379;  
XX  
XX 21-OCT-2002 (first entry)  
XX  
XX ScbA, ScbR and ScbB encoding genes containing DNA sequence.  
XX  
XX Antibiotic; bacterium; scbA; afsA; scbR; arpa; barA; actinorhodine; Act;  
XX KW undecylprodigiosin; Red; gene; ds.  
XX  
XX Streptomyces coelicolor.  
XX  
XX CA2322241-A1.  
XX  
XX 23-APR-2002.  
XX  
XX 23-OCT-2000; 2000CA-02322241.  
XX  
XX 23-OCT-2000; 2000CA-02322241.  
XX  
XX (PLAN-) PLANT BIOSCIENCE LTD.  
XX  
XX Takano E, Bibb M;  
XX  
XX WPI; 2002-501089/54.  
XX DR P-PSDB; ABB80940, ABB80941, ABB80942.  
XX  
XX Modifying antibiotic-producing Streptomyces, to increase, or alter timing  
XX of, antibiotic production, by deleting the scbA or scbR genes.  
XX  
XX Claim 19, 20; Fig 14; 64pp; English.  
XX  
XX The invention provides a method for modifying an antibiotic-producing
```

```
CC strain of Streptomyces to increase production of antibiotics or to alter  
CC the timing of antibiotic production. The modification is functional  
CC deletion of the scbA gene of S. coelicolor, or its homologues, but is not  
CC deletion of the afsA gene of S. griseus, or the modification is  
CC functional deletion of the scbR gene of S. coelicolor, or its homologues,  
CC but is not deletion of arpa of S. griseus nor barA of S. virginiae. The  
CC method is particularly used for production of the antibiotics  
CC actinorhodine (Act) and undecylprodigiosin (Red). The present sequence  
CC represents the nucleic acid sequence of the region containing the S.  
CC coelicolor ScbA, ScbR and ScbB proteins  
XX  
XX Sequence 4346 BP; 650 A; 1561 C; 1519 G; 616 T; 0 U; 0 Other;  
XX  
XX Query Match 80.0%; Score 19.2; DB 6; Length 4346;  
XX Best Local Similarity 75.0%; Pred. No. 3,7e+02;  
XX Matches 18; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
XX  
XX  
XX 1 GTCTGTGCGCGTSGACGCGSAC 24  
XX 1250 GTCTGTGCGCGGTGACCGCAC 1273  
XX  
XX Db  
XX  
XX RESULT 4  
XX ADE14794  
XX ID ADE14794 standard; DNA; 4346 BP.  
XX  
XX ADE14794;  
XX  
XX 29-JAN-2004 (first entry)  
XX  
XX Streptomyces coelicolor partial genome sequence 2.  
XX  
XX Streptomyces coelicolor antibiotic production; scbA gene; ScbR gene;  
XX KW actinorhodin; undecylprodigiosin; gene; ds.  
XX  
XX Streptomyces coelicolor.  
XX  
XX  
XX Key Location/Qualifiers  
XX FH complement(1199..2143)  
XX FT /tag= a  
XX FT /product= "Streptomyces coelicolor ScbB protein"  
XX FT /transl_except= (3149..3147, aa:Tyr)  
XX FT 2261..2908  
XX FT /tag= b  
XX FT /product= "Streptomyces coelicolor ScbR protein"  
XX FT complement(3024..3812)  
XX FT /tag= c  
XX FT /product= "Streptomyces coelicolor ScbA protein"  
XX FT /transl_except= (3150..3145, aa:Met)  
XX  
XX US2003124644-A1.  
XX  
XX 03-JUL-2003.  
XX  
XX 23-OCT-2001; 2001US-00017471.  
XX  
XX 23-OCT-2000; 2000US-0242561P.  
XX  
XX (TAKA/) TAKANO E.  
XX PA (BIBB/) BIBB M J.  
XX  
XX Takano E, Bibb MJ;  
XX  
XX WPI; 2003-810983/76.  
XX DR P-PSDB; ADE14791, ADE14792, ADE14793.  
XX  
XX Modifying an antibiotic-producing strain of Streptomyces coelicolor or  
XX Streptomyces lividans to increase or alter the timing of antibiotic  
XX production in the strain, comprises functionally deleting in the strain  
XX the scbA or ScbR gene.  
XX  
XX Claim 19; Fig 14; 33pp; English.  
XX
```

CC This invention relates to the novel modification of an antibiotic-  
CC producing strain of Streptomyces coelicolor or Streptomyces lividans to  
CC increase or to alter the timing of antibiotic production in the strain.  
CC The method comprises functionally deleting in the strain the scdA or scbR  
CC gene. The method is useful in increasing and altering the timing of  
CC antibiotic production (especially actinorhodin and undecylprodigiosin) in  
CC Streptomyces species, particularly Streptomyces coelicolor or  
CC Streptomyces lividans. The present sequence is that of a region of the  
CC Streptomyces coelicolor genome, which encodes the scdA, scbB and scbR  
CC proteins and which is related to the invention.  
XX  
SQ Sequence 4346 BP, 650 A; 1561 C; 1519 G; 616 T; 0 U; 0 Other;  
Query Match 80.0%; Score 19.2; DB 10; Length 4346;  
Best Local Similarity 75.0%; Pred. No. 3.7e+02;  
Matches 18; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
Gy 1 GTCCTGSGCCSGTSACSGSAC 24  
Db 1250 GTCCTGATGCCCGTGACCCGCAC 1273  
RESULT 5  
ABD11456  
ID ABD11456 standard; DNA; 669 BP.  
XX  
AC ABD11456;  
XX  
XX 29-JUL-2004 (first entry)  
XX  
DB Pseudomonas aeruginosa polynucleotide #10060.  
XX  
KM Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
KW antibacterial.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN US6551795-B1.  
XX  
PD 22-APR-2003.  
XX  
PF 18-FEB-1999; 99US-00252991.  
XX  
PR 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX  
XX WPI: 2003-615309/58.  
DR P-PSDB; ABO77885.  
XX  
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
XX Disclosure; SEQ ID NO.10060; 455bp; English.  
XX  
CC The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biochip technology. Sequences ABD01397-  
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed

CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html  
XX  
SQ Sequence 669 BP, 80 A; 230 C; 229 G; 130 T; 0 U; 0 Other;  
Query Match 78.3%; Score 18.8; DB 11; Length 669;  
Best Local Similarity 70.8%; Pred. No. 5.4e+02;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
Gy 1 GTCCTGSGCCSGTSACSGSAC 24  
Db 252 GTCCTGCTTGCCGCGCACCCGCAC 275  
RESULT 6  
ABD11169/c  
ID ABD11169 standard; DNA; 810 BP.  
XX  
AC ABD11169;  
XX  
XX 29-JUL-2004 (first entry)  
XX  
DB Pseudomonas aeruginosa polynucleotide #9773.  
XX  
KM Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
KW antibacterial.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN US6551795-B1.  
XX  
PD 22-APR-2003.  
XX  
PF 18-FEB-1999; 99US-00252991.  
XX  
PR 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX  
XX WPI: 2003-615309/58.  
DR P-PSDB; ABO77598.  
XX  
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
XX Disclosure; SEQ ID NO.9773; 455bp; English.  
XX  
CC The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biochip technology. Sequences ABD01397-  
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html  
XX  
SQ Sequence 810 BP, 155 A; 278 C; 275 G; 102 T; 0 U; 0 Other;  
Query Match 78.3%; Score 18.8; DB 11; Length 810;  
Best Local Similarity 70.8%; Pred. No. 5.4e+02;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;





CC content, improved yield by modification of carbohydrate, nitrogen or  
CC phosphorus use and/or uptake, by modification of photosynthesis or by  
CC providing improved plant growth and development under at least one stress  
CC condition, improved lignin production or improved galactomannan  
CC production. This sequence represents a bacterial polynucleotide used in  
CC the scope of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification but was obtained in electronic  
CC format from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
SQ Sequence 1136 BP; 223 A; 407 C; 328 G; 178 T; 0 U; 0 Other;  
XX  
Query Match 78.3%; Score 18.8; DB 13; Length 1136;  
Best Local Similarity 70.8%; Pred. No. 5.4e+02;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
Gy 1 GTCCGTGCGCCGCGTACGCGSAC 24  
Db 639 GTCCGTGCGTCCGCGTACGCGAC 616  
RESULT 9  
ID ADS63742 standard; cDNA; 1136 BP.  
AC ADS63742;  
XX  
XX 02-DEC-2004 (first entry)  
XX  
XX Bacterial polynucleotide #15729.  
XX  
XX Recombinant DNA construct; transformed plant; improved plant property;  
XX cold tolerance; heat tolerance; drought tolerance; herbicide; osmotic;  
XX pathogen tolerance; pest tolerance; plant disease resistance;  
XX cell cycle pathway modification; plant growth regulator;  
XX homologous recombination; seed oil yield; protein yield; carbohydrate;  
XX nitrogen; phosphorus; photosynthesis; lignin; galactomannan;  
XX bacterial polynucleotide; gene; ss.  
XX  
XX Bacteria.  
XX  
XX US2003233675-A1.  
XX  
XX 18-DEC-2003.  
XX  
XX 20-FEB-2003; 2003US-00369493.  
XX  
XX 21-FEB-2002; 2002US-0360039P.  
XX  
XX (CAOY/) CAO Y.  
XX (HINK/) HINKLE G J.  
XX (SLAT/) SLATER S C.  
XX (CHEN/) CHEN X.  
XX (GOLD/) GOLDMAN B S.  
XX  
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;  
XX WPI; 2004-061375/06.  
XX  
XX New recombinant DNA construct comprising a promoter positioned to provide  
XX for expression of a polynucleotide encoding a polypeptide from a  
XX microbial source, useful for producing plants with improved properties.  
XX  
XX Claim 1; SEQ ID NO 39416; 122pp; English.  
XX  
XX The invention relates to a recombinant DNA construct comprising a  
XX promoter functional in a plant cell, where the promoter is positioned to  
XX provide for expression of a polynucleotide encoding a polypeptide from a  
XX microbial source. The invention also relates to a transformed plant  
XX comprising the recombinant DNA construct and a method of producing a  
XX transformed plant having an improved property. The plant is a crop plant  
XX such as maize or soybean. The method of producing a transformed plant  
XX having an improved property comprises transforming a plant with the  
XX recombinant DNA construct and growing the transformed plant, where the

CC polynucleotide or polypeptide is useful for improving plant properties.  
CC The recombinant DNA construct is useful for producing plants with  
CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
CC increased resistance to plant disease, better growth rate by modification  
CC of the cell cycle pathway with plant growth regulators, increased rate of  
CC homologous recombination, modified seed oil or protein yield and/or  
CC content, improved yield by modification of carbohydrate, nitrogen or  
CC phosphorus use and/or uptake, by modification of photosynthesis or by  
CC providing improved plant growth and development under at least one stress  
CC condition, improved lignin production or improved galactomannan  
CC production. This sequence represents a bacterial polynucleotide used in  
CC the scope of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification but was obtained in electronic  
CC format from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
SQ Sequence 1136 BP; 223 A; 407 C; 328 G; 178 T; 0 U; 0 Other;  
XX  
Query Match 78.3%; Score 18.8; DB 13; Length 1136;  
Best Local Similarity 70.8%; Pred. No. 5.4e+02;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
Gy 1 GTCCGTGCGCCGCGTACGCGSAC 24  
Db 639 GTCCGTGCGTCCGCGTACGCGAC 616  
RESULT 10  
ID ABD09558/c  
XX ABD09558 standard; DNA; 1338 BP.  
XX  
XX ABD09558;  
XX  
XX 29-JUL-2004 (first entry)  
XX  
XX Pseudomonas aeruginosa polynucleotide #8162.  
XX  
XX  
XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
XX antibacterial.  
XX  
XX Pseudomonas aeruginosa.  
XX  
XX US651795-B1.  
XX  
XX 22-Apr-2003.  
XX  
XX 18-FEB-1999; 99US-00252991.  
XX  
XX 18-FEB-1998; 98US-0074788P.  
XX 27-JUL-1998; 98US-0094190P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX  
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX WPI; 2003-615309/58.  
XX  
XX P-PsDB; ABO75987.  
XX  
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
XX useful as molecular targets for diagnostics, prophylaxis and treatment of  
XX pathological conditions resulting from bacterial infection.  
XX  
XX Disclosure; SEQ ID NO 8162; 455pp; English.  
XX  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
XX polynucleotides encoding them. The sequences are useful in diagnosis and  
XX therapy of pathological conditions, as molecular targets for diagnostics,  
XX prophylaxis and treatment of pathological conditions resulting from a  
XX bacterial infection, for evaluating a compound, such as a polypeptide,  
XX for the ability to bind a P. aeruginosa nucleic acid, as components of  
XX effective antibacterial targets, as targets for antibacterial drugs,  
XX including anti-P. aeruginosa drugs, as templates for recombinant  
XX production of P. aeruginosa-derived peptides or polypeptides, as target

CC components for diagnosis and/or treatment of *P. aeruginosa*-caused  
 CC infection, and in detection of *P. aeruginosa* sequences or other sequences  
 CC of *Pseudomonas* species using biochip technology. Sequences ABD01397-  
 CC ABD17967 represent *P. aeruginosa* polynucleotides of the invention. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format from USPTO at  
 CC [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html)  
 CC  
 XX  
 XX Sequence 1338 BP, 333 A, 404 C, 347 G, 254 T, 0 U, 0 Other;  
 SQ  
 Query Match 78.3%; Score 18.8; DB 11; Length 1338;  
 Best Local Similarity 70.8%; Pred No. 5.4e+02;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0  
 Oy 1 GTCTGTGTCGCCGTGACGCGSAC 24  
 |||||:|||||:|||||:|||||:  
 Db 999 GTCCGCTGTGCCGTGACCCGAC 976  
 RESULT 11  
 ABD11263/c  
 ID ABD11263 standard; DNA; 4248 BP.  
 XX  
 XX ABD11263;  
 AC  
 XX 29-JUL-2004 (first entry)  
 DT  
 XX  
 XX *Pseudomonas aeruginosa* polynucleotide #9867.  
 DE  
 XX  
 XX Bacterial infection; gene; ds; *Pseudomonas aeruginosa* infection;  
 KM antibacterial.  
 KW  
 XX  
 XX *Pseudomonas aeruginosa*.  
 OS  
 XX  
 XX US6551795-B1.  
 PN  
 XX  
 XX 22-APR-2003.  
 PD  
 XX  
 XX 18-FEB-1999; 99US-00252991.  
 PF  
 XX  
 XX 18-FEB-1998; 98US-0074788P.  
 PR  
 XX  
 XX 27-JUL-1998; 98US-0094190P.  
 PR  
 XX  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 PA  
 XX  
 XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
 PI  
 XX  
 XX WPI; 2003-615309/58.  
 DR P-PSDB; ABD07692.  
 XX  
 XX Novel isolated nucleic acid encoding *Pseudomonas aeruginosa* polypeptide,  
 PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
 PT pathological conditions resulting from bacterial infection.  
 XX  
 XX  
 XX Disclosure; SEQ ID NO 9867; 455BP; English.  
 PS  
 XX  
 XX The invention relates to *Pseudomonas aeruginosa* polypeptides and the  
 CC polynucleotides encoding them. The sequences are useful in diagnosis and  
 CC therapy of pathological conditions, as molecular targets for diagnostics,  
 CC prophylaxis and treatment of pathological conditions resulting from a  
 CC bacterial infection, for evaluating a compound, such as a polypeptide,  
 CC for the ability to bind a *P. aeruginosa* nucleic acid, as components of  
 CC effective antibacterial targets, as targets for antibacterial drugs,  
 CC including anti-*P. aeruginosa* drugs, as templates for recombinant  
 CC production of *P. aeruginosa*-derived peptides or polypeptides, as target  
 CC components for diagnosis and/or treatment of *P. aeruginosa*-caused  
 CC infection, and in detection of *P. aeruginosa* sequences or other sequences  
 CC of *Pseudomonas* species using biochip technology. Sequences ABD01397-  
 CC ABD17967 represent *P. aeruginosa* polynucleotides of the invention. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format from USPTO at  
 CC [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html)  
 CC  
 XX

SQ Sequence 4248 BP; 832 A; 1408 C; 1389 G; 619 T; 0 U; 0 Other;  
 Query Match 78.3%; Score 18.8; DB 11; Length 4248;  
 Best Local Similarity 70.8%; Pred. No. 5.3e+02;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 GTCTGTGTGCGCGGTACSCGAC 24  
 |||||:||||:||||:||||  
 DB 975 GTCTGTGTGCGCGCGCACCCGAC 952  
 RESULT 12  
 ID ACA42624/c  
 AC ACA42624 standard; DNA; 4254 BP.  
 AC ACA42624;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Prokaryotic essential gene #24281.  
 XX  
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 KW drug design; gene.  
 XX  
 OS Pseudomonas aeruginosa.  
 OS  
 PN WO200271183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 XX  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 PA  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX  
 DR WPI; 2003-029926/02.  
 DR P-PSDB; ABU38754.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation or  
 PT isolate candidate molecules for rational drug discovery programs.  
 PS  
 PS Claim 14; SEQ ID NO 30494; 1766bp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-regulated gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism's activity; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the

proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/publ/published\\_pct\\_sequences](http://wipo.int/pub/publ/published_pct_sequences)

Sequence 4254 BP; 837 A; 1404 C; 1389 G; 624 T; 0 U; 0 Other;

Query Match 78.3%; Score 18.8; DB 8; Length 4254;  
Best Local Similarity 70.8%; Pred. No. 5.3e+02;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

1 GTCCGTGCGCGGTGACGCGSAC 24  
927 GTCTGCTTGGCGGCGACCGCGAC 904

RESULT 13  
AAT80415  
ID AAT80415 standard; DNA; 13987 BP.  
XX  
AC AAT80415;  
XX  
DT 02-MAR-1998 (first entry)  
XX  
DE Hybrid smgG/tylG ORF1.  
XX  
KW Ty lactone synthase gene cluster; tylG gene; multifunctional protein;  
KM platenolide synthase gene cluster; platenolide production; smg gene;  
XX polyketide; ty lactone synthesis; antibiotic; tylosin; hybrid gene; ss.  
OS Streptomyces ambofaciens.  
XX  
FH Streptomyces fradiae.  
XX  
FH Key Location/Qualifiers  
FT CDS 350..13987  
FT /tag= a  
FT /transl\_except= (pos:350..352, aa:Met)  
FT /note= "ORF1 encodes hybrid protein shown in AAW22611"

EP791655-A2.  
27-AUG-1997.  
XX  
PD 27-AUG-1997.  
XX  
PF 19-FEB-1997; 97EP-00301056.  
XX  
PR 22-FEB-1996; 96US-0012078P.  
XX  
PA (EILIL ) LILLY & CO ELI.  
PI Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;  
XX WPI: 1997-418046/39.  
DR P-PSDB; AAW22611.  
XX  
PT DNA encoding Streptomyces fradiae ty lactone synthase domain - for  
XX production of tylosin-related polyketide compounds.  
XX  
PS Claim 22; Page 178-197; 220pp; English.

This sequence represents a hybrid gene of the invention. This sequence was created by replacing a EcoRI-ApaI fragment of smg ORF1 with a EcoRI-SmaI fragment from tylG ORF1. The position of the nucleotides from each of the two genes is not given in the specification. The smg gene (see AAT80414) was isolated from Streptomyces ambofaciens, and encodes the multi-functional proteins which direct the synthesis of the polyketide platenolide. Platenolide is the basic building block of the macroide antibiotic epiramyacin. The tylG gene (see AAT80413) is the ty lactone

synthase gene cluster of the invention. The tylG sequence was isolated from Streptomyces fradiae, and encodes multifunctional proteins which direct the synthesis of the polyketide ty lactone. Ty lactone is the basic building block of the antibiotic tylosin. The hybrid sequence can be used to transform *S. ambofaciens* lacking the smg ORF1 sequence, or *S. fradiae* lacking the tylG ORF1 sequence, so that they can produce polyketides. The DNA sequence can be modified so as to alter the type of carboxylic acids incorporated, the number of carboxylic acids incorporated and/or the post-condensation reactions performed, thereby resulting in novel tylosin-related polyketides

Sequence 13987 BP; 1556 A; 4401 C; 5727 G; 2303 T; 0 U; 0 Other;

Query Match 78.3%; Score 18.8; DB 2; Length 13987;  
Best Local Similarity 70.8%; Pred. No. 5.2e+02;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

1 GTCCGTGCGCGGTGACGCGSAC 24  
8324 GACCTGGTGGCGGTGACGCGCAC 8347

RESULT 14  
AAT80414  
ID AAT80414 standard; DNA; 44377 BP.  
XX  
AC AAT80414;  
XX  
DT 27-FEB-1998 (first entry)  
XX  
DE Platenolide synthase gene cluster.  
XX  
KW Ty lactone synthase gene cluster; tylG gene; multifunctional protein;  
KM platenolide synthase gene cluster; platenolide production; smg gene;  
XX polyketide; ty lactone synthesis; antibiotic; tylosin; ss.  
OS Streptomyces ambofaciens.  
XX  
FH Streptomyces fradiae.  
XX  
FH Key Location/Qualifiers  
FT CDS 350..14002  
FT /tag= a  
FT /transl\_except= (pos:350..352, aa:Met)  
FT /note= "ORF1 encodes protein shown in AAW22606"  
FT 14046..20036  
FT /tag= b  
FT /note= "ORF2 encodes protein shown in AAW22607"  
FT 20110..31284  
FT /tag= c  
FT /transl\_except= (pos:20111..20113, aa:Met)  
FT /note= "ORF3 encodes protein shown in AAW22608"  
FT 31329..36071  
FT /tag= d  
FT /note= "ORF4 encodes protein shown in AAW22609"  
FT 36155..41830  
FT /tag= e  
FT /note= "ORF5 encodes protein shown in AAW22610"

EP791655-A2.  
27-AUG-1997.  
XX  
PD 27-AUG-1997.  
XX  
PF 19-FEB-1997; 97EP-00301056.  
XX  
PR 22-FEB-1996; 96US-0012078P.  
XX  
PA (EILIL ) LILLY & CO ELI.  
PI Dehoff BS, Kuhstoss SA, Rosteck PR, Sutton KL;  
XX WPI: 1997-418046/39.  
DR P-PSDB; AAW22606, AAW22607, AAW22608, AAW22609, AAW22610.  
XX  
PT DNA encoding Streptomyces fradiae ty lactone synthase domain - for

Query Match	Best Local Similarity	Matches	17, Conservative	5, Mismatches	2, Indels	0, Gaps	0, Mismatches
QY	1						
Db	8339	GACCTGTCGCGCGTACCGGAC	8362				
RESULT 15							
ID	AAT78508	standard, DNA; 44377 BP.					
XX	AAT78508;						
XX	26-FEB-1998	(first entry)					
DT	Platenolide synthase gene cluster.						
XX	Platenolide synthase gene cluster.						
XX	Platenolide synthase gene cluster; platenolide production; smg gene;						
KW	multi-functional protein; macrolide antibiotic; spiramycin; ss.						
XX	Streptomyces ambofaciens.						
OS	Streptomyces ambofaciens.						
XX	Key	Location/Qualifiers					
XX	CDS	350..14002					
FT		/*tag= a					
FT		/transl_except= (pos:350..352, aa:Met)					
FT		/note= "ORF1 encodes protein shown in AAW23716"					
FT	CDS	14046..20036					
FT		/*tag= b					
FT		/note= "ORF2 encodes protein shown in AAW23717"					
FT	CDS	20110..31284					
FT		/*tag= c					
FT		/transl_except= (pos:20111..20113, aa:Met)					
FT		/note= "ORF3 encodes protein shown in AAW23718"					
FT	CDS	31329..36071					
FT		/*tag= d					
FT		/note= "ORF4 encodes protein shown in AAW23719"					
FT	CDS	36155..41830					
FT		/*tag= e					
FT		/note= "ORF5 encodes protein shown in AAW23720"					
XX	BP791656-A2.						
XX	27-AUG-1997.						
PD	19-FEB-1997;	97EP-00301066.					

XX	22-FEB-1996;	96US-0012050P.
PR	(ELIL ) LILLY & CO ELI.	
XX	Burgett SG, Kuhseas SA, Rao RN, Richardson MA, Rosteck PR;	
XX	WPI, 1997-418047/39.	
XX	P-PSDB; AAM23716, AAM23717, AAM23718, AAM23719, AAM23720.	
XX	DNA encoding Streptomyces ambifaciens plateanoliide synthase domain - for	
XX	production of spiramycin-related polyketide antibiotics.	
XX	Claim 9; Page 8-33; 81pp; English.	
XX	This sequence represents the plateanoliide synthase gene cluster of the	
XX	invention. This sequence is referred to as the smg gene, and was	
XX	isolated from Streptomyces ambifaciens. This sequence encodes the multi-	
XX	functional proteins which direct the synthesis of the polyketide	
XX	plateanoliide. Plateanoliide is the basic building block of the macroliide	
XX	antibiotic spiramycin. The DNA can be used to produce compounds	
XX	exhibiting antibiotic activity based on the plateanoliide structure,	
XX	including specifically the macroliide antibiotic spiramycin and spiramycin	
XX	analogues and derivatives. Modifications of the plateanoliide synthase DNA	
XX	sequence can be made so as to change the number and type of carboxylic	
XX	acids incorporated into the growing polyketide chain and to change the	
XX	kind of post-condensation processing that is conducted	
XX	Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 U; 0 Other;	
XX	Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 U; 0 Other;	
XX	Query Match 78.3%; Score 18.8; DB 2; Length 44377;	
XX	Best Local Similarity 70.8%; Pred. No. 5.1e+02;	
XX	Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0	
OY	1 GTCTGTGTGCGCCGTGACGCGSAC 24	
DB	:     :     :	
	8339 GACCTGTGTGCGCCGTGACGCGGAC 8362	
RESULT 16		
ADA71299/C		
ID	ADA71299 standard; DNA; 3168 BP.	
XX	ADA71299;	
XX	20-NOV-2003 (first entry)	
XX	Rice gene, SEQ ID 4622.	
XX	Plant; bacterial infection; fungal infection; viral infection; rice;	
XX	gene; ds.	
XX	Oryza sativa.	
XX	WO2003000898-A1.	
XX	03-JAN-2003.	
XX	22-JUN-2001; 2001WO-IB001105.	
XX	22-JUN-2001; 2001WO-IB001105.	
XX	(SYGN ) SYNGENTA PARTICIPATIONS AG.	
XX	Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;	
XX	Katagiri F, Qian S, Tao Y, Whitam S, Xie Z, Zhu T, Zou G;	
XX	WPI; 2003-175290/17.	
XX	Identifying at least one gene involved in plant resistance or response to	
XX	pathogenic infection for conferring resistance or tolerance to a plant to	
XX	bacterial, fungal or viral infection by determining or detecting plant	
XX	gene expression.	





CC containing the frenolicin gene cluster DNA sequence operably connected to  
CC an expression control sequence; (2) a host cell transformed by the above  
CC vector; (3) a protein coded by the above DNA sequence; (4) a method for  
CC the preparation of frenolicin or a biosynthetic intermediate for it in  
CC which the above cell is cultured and frenolicin or its biosynthetic  
CC intermediate is isolated from the culture or the cell; (5) a method for  
CC the preparation of frenolicin B by oxidizing frenolicin, and (6) a method  
CC for the preparation of a feed composition by mixing frenolicin with other  
CC components. Frenolicin B is useful as an antibiotic  
XX

SO Sequence 24379 BP; 3077 A; 9792 C; 8499 G; 3011 T; 0 U; 0 Other;

Query Match 75.0%; Score 18; DB 2; Length 24379;  
Best Local Similarity 75.0%; Pred. No. 1.1e+03;  
Matches 15; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 5 TGCTGGCCGCTGACCGGAC 24  
|||:|||||:|||||:|||||:  
DB 7005 TGCTGGCCGCTGACCGGAC 7024

RESULT 19  
ABD02704/C  
ID ABD02704 standard; DNA; 375 BP.  
XX  
AC ABD02704;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Pseudomonas aeruginosa polynucleotide #1308.  
XX  
KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
KW antibacterial.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN US6551795-B1.  
XX  
PD 22-APR-2003.  
XX  
PF 18-FEB-1999; 99US-00252991.  
XX  
PR 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX  
DR WPI; 2003-615309/58.  
DR P-PSDB; ABO69133.  
XX  
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
PS Disclosure; SEQ ID NO 1308; 455pp; English.  
XX  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biochip technology. Sequences ABD01397-  
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at

CC seqdata.uspto.gov/sequence.html  
XX  
SQ Sequence 375 BP; 74 A; 130 C; 138 G; 33 T; 0 U; 0 Other;

Query Match 74.2%; Score 17.8; DB 11; Length 375;  
Best Local Similarity 69.6%; Pred. No. 1.4e+03;  
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTCTGGCCGCTGACCGGGA 23  
|||:|||||:|||||:|||||:  
DB 117 GTCTGGCCGCTGACCGGGA 95

RESULT 20  
ABD02553  
ID ABD02553 standard; DNA; 681 BP.  
XX  
AC ABD02553;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Pseudomonas aeruginosa polynucleotide #1157.  
XX  
KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
KW antibacterial.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN US6551795-B1.  
XX  
PD 22-APR-2003.  
XX  
PF 18-FEB-1999; 99US-00252991.  
XX  
PR 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX  
DR WPI; 2003-615309/58.  
DR P-PSDB; ABO68982.  
XX  
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
PS Disclosure; SEQ ID NO 1157; 455pp; English.  
XX  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biochip technology. Sequences ABD01397-  
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html  
XX

Query Match 74.2%; Score 17.8; DB 11; Length 681;  
Best Local Similarity 69.6%; Pred. No. 1.3e+03;  
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;



OY 1 GTCCTGCTGACCGGTGACCGGA 23  
DB 610 GTCCTGCTGACCGGTGACCGGA 632

RESULT 21  
ADP04866  
ID ADP04866 standard; cDNA, 1455 BP.  
XX  
XX ADP04866;  
AC  
XX  
XX 29-JUL-2004 (first entry)  
DT  
XX  
DE Sea squirt cDNA with tissue specific expression in development Seq 461.  
XX  
XX gene; ss; sea squirt; regeneration medicine; gene therapy;  
XX  
XX cell proliferation; differentiation; reproduction;  
XX  
XX environmental measurement; water survey.  
OS  
XX Ciona intestinalis.  
XX  
XX JP2004057129-A.  
XX  
XX 26-FEB-2004.  
PD  
XX  
XX 31-JUL-2002; 2002JP-00222593.  
PF  
XX  
XX 31-JUL-2002; 2002JP-00222593.  
PR  
XX  
XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.  
XX  
XX WPI, 2004-287079/27.  
DR  
XX P-PSDB; ADP04867.  
XX  
XX Novel gene cluster which is specifically expressed in tissue or organ  
PT during development phase of sea squirt, useful for elucidation of  
PT mechanism of development of tissue or organ of sea squirt.  
XX  
XX Claim 2; SEQ ID NO 461; 1846bp; Japanese.  
PS  
XX  
XX This invention relates to novel genes and the encoded proteins thereof  
CC that are derived from the sea squirt Ciona intestinalis. Specifically, it  
CC refers to those genes that are expressed in the tissues or organs of the  
CC sea squirt during its developmental phase. The present invention  
CC describes the identification of these genes as useful for elucidation of  
CC the mechanism of development and hence for developing regeneration  
CC medicines and gene therapy techniques. Accordingly, they can be used in  
CC the research of various genetic diseases, as well as the analysis of cell  
CC proliferation, differentiation and reproduction. Furthermore, such  
CC compositions can be useful for environmental measurements and water  
CC surveys, particularly for sea water surveys, and also for the preparation  
CC of transformed sea squirt for improving edibility of sea squirt such as  
CC Halocynthia roretzi. This polynucleotide sequence is a sea squirt cDNA  
CC sequence that exhibits tissue specific expression during development,  
CC given in an exemplification of the invention.  
XX  
XX SQ Sequence 1455 BP; 412 A; 258 C; 364 G; 421 T; 0 U; 0 Other;

Query Match 74.2%; Score 17.8; DB 12; Length 1455;  
Best Local Similarity 69.6%; Pred. No. 1.3e+03;  
Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTCCTGCTGACCGGTGACCGGA 23  
DB 984 GTTCTTGTGCGCGGTGACCGGA 1006

RESULT 22  
ABD07378  
ID ABD07378 standard; DNA; 492 BP.  
XX  
XX ABD07378;  
XX

DT 29-JUL-2004 (first entry)  
XX  
DE Pseudomonas aeruginosa polynucleotide #5982.  
XX  
XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
XX  
XX antibacterial.  
XX  
XX Pseudomonas aeruginosa.  
OS  
XX US5551795-B1.  
XX  
XX 22-APR-2003.  
PD  
XX  
XX 18-FEB-1999; 99US-00252991.  
PF  
XX  
XX 18-FEB-1998; 98US-0074788P.  
PR  
XX 27-JUL-1998; 98US-0094190P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX  
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX  
XX WPI, 2003-615309/58.  
DR  
XX P-PSDB; ABO73807.  
XX  
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
XX Disclosure; SEQ ID NO 5982; 455bp; English.  
PS  
XX  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biochip technology. Sequences ABD01397-  
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC segata.uspto.gov/sequence.html  
XX  
XX SQ Sequence 492 BP; 57 A; 174 C; 162 G; 99 T; 0 U; 0 Other;

Query Match 72.5%; Score 17.4; DB 11; Length 492;  
Best Local Similarity 71.4%; Pred. No. 1.9e+03;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 2 TCCCTGCTGACCGGTGACCGG 22  
DB 232 TCCCTGCTGACCGGTGACCGG 252

RESULT 23  
ABD01735/C  
ID ABD01735 standard; DNA; 654 BP.  
XX  
XX ABD01735;  
XX  
XX 29-JUL-2004 (first entry)  
DT  
XX  
XX Pseudomonas aeruginosa polynucleotide #339.  
DE  
XX  
XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
XX  
XX antibacterial.  
XX  
XX Pseudomonas aeruginosa.  
OS



XX US6551795-B1.  
PN 22-APR-2003.  
XX  
XX 18-FEB-1999; 99US-00252991.  
PF  
XX 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
XX WPI; 2003-615309/58.  
DR P-PSDB; ABO68164.  
XX  
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
PS Disclosure; SEQ ID NO 339; 455bp; English.  
XX  
CC The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using bioclip technology. Sequences ABD01397-  
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html  
XX  
SQ Sequence 654 BP; 89 A; 214 C; 259 G; 92 T; 0 U; 0 Other;  
Query Match 72.5%; Score 17.4; DB 11; Length 654;  
Best Local Similarity 71.4%; Pred. No. 1.9e+03;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
QY 2 TCGTGTGGCCGCGTACGCGS 22  
DB 548 TCGTGTGGCCGCGTACGCGG 528  
RESULT 24  
AAS54125  
ID AAS54125 standard; DNA; 744 BP.  
XX  
XX AAS54125;  
AC  
XX 13-FEB-2002 (first entry)  
DT  
XX Pseudomonas aeruginosa DNA for cellular proliferation protein #256.  
DE  
XX Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;  
KW antibacterial; drug design.  
XX  
XX Pseudomonas aeruginosa.  
OS  
XX MO200170955-A2.  
PN  
XX 27-SEP-2001.  
PD  
XX 21-MAR-2001; 2001WO-US009180.  
PF  
XX 21-MAR-2000; 2000US-0191078P.  
PR

PR 23-MAY-2000; 2000US-0206848P.  
PR 26-MAY-2000; 2000US-0207727P.  
PR 23-OCT-2000; 2000US-0242578P.  
PR 27-NOV-2000; 2000US-0253625P.  
PR 22-DEC-2000; 2000US-0257931P.  
PR 16-FEB-2001; 2001US-0269308P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
PA  
PI Haseelbeck R, Ohlsen KL, Zykkind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
XX WPI; 2001-611495/70.  
DR P-PSDB; AAU36266.  
XX  
XX New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids.  
PT  
XX  
XX Claim 27; SEQ ID NO 7762; 511bp; English.  
PS  
CC The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the genes,  
CC their use in the discovery of novel antibiotics, the essential genes  
CC themselves and the encoded proteins. The prokaryotes used are Escherichia  
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,  
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also  
CC useful for the identification of potential new targets for antibiotic  
CC development. The antisense nucleic acids can also be used to identify  
CC proteins used in proliferation, to express these proteins, and to obtain  
CC antibodies capable of binding to the expressed proteins. The proteins can  
CC be used to screen compounds in rational drug discovery programmes. The  
CC antisense nucleic acid sequence is also useful to screen for homologous  
CC nucleic acids which are required for cell proliferation in a wide variety  
CC of organisms. The present sequence encodes an essential prokaryotic  
CC cellular proliferation protein. Note: The sequence data for this patent  
CC did not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 744 BP; 103 A; 281 C; 257 G; 103 T; 0 U; 0 Other;  
Query Match 72.5%; Score 17.4; DB 4; Length 744;  
Best Local Similarity 71.4%; Pred. No. 1.9e+03;  
Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
QY 2 TCGTGTGGCCGCGTACGCGS 22  
DB 602 TCGTGTGGCCGCGTACGCGG 622  
RESULT 25  
ACA42298  
ID ACA42298 standard; DNA; 744 BP.  
XX  
XX ACA42298;  
AC  
XX 19-JUN-2003 (first entry)  
DT  
XX Prokaryotic essential gene #23955.  
DE  
XX Antisense; ds; prokaryotic essential gene; cell proliferation;  
KW drug design; gene.  
XX  
XX Pseudomonas aeruginosa.  
OS  
XX MO200277183-A2.  
PN  
XX 03-OCT-2002.  
PD  
XX 21-MAR-2002; 2002WO-US009107.  
PF  
XX 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR

25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
DR P-PSDB; ABU38428.  
XX  
XX New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation or  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 14; SEQ ID NO 30168; 1766pp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
CC prokaryotic essential genes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 744 BP; 103 A; 281 C; 257 G; 103 T; 0 U; 0 Other;  
SQ  
XX  
XX Query Match 72.5%; Score 17.4; DB 8; Length 744;  
XX Best Local Similarity 71.4%; Pred. No. 1.9e+03;  
XX Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
OY 2 TCCTGCTGCGCCGTCGACGCGS 22  
DB 602 TCCTGCTGCGCCGTCGACGCGS 622

XX  
XX Pseudomonas aeruginosa.  
OS  
XX US6551795-B1.  
XX  
XX 22-APR-2003.  
XX  
XX 18-FEB-1999; 99US-00252991.  
XX  
XX 18-FEB-1998; 98US-0074788P.  
PR 27-JUL-1998; 98US-0094190P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX  
XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;  
PI WPI; 2003-615309/58.  
DR P-PSDB; ABO68155.  
XX  
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
PT pathological conditions resulting from bacterial infection.  
XX  
XX Disclosure; SEQ ID NO 330; 455pp; English.  
XX  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
CC polynucleotides encoding them. The sequences are useful in diagnosis and  
CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biochip technology. Sequences ABD01397-  
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html  
XX  
XX Sequence 759 BP; 108 A; 283 C; 262 G; 106 T; 0 U; 0 Other;  
SQ  
XX  
XX Query Match 72.5%; Score 17.4; DB 11; Length 759;  
XX Best Local Similarity 71.4%; Pred. No. 1.9e+03;  
XX Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
OY 2 TCCTGCTGCGCCGTCGACGCGS 22  
DB 617 TCCTGCTGCGCCGTCGACGCGS 637

RESULT 27  
ACH97630  
ID ACH97630 standard; DNA; 843 BP.  
XX  
XX ACH97630;  
XX  
XX 29-JUL-2004 (first entry)  
XX  
XX Klebsiella pneumoniae polynucleotide seqid 3425.  
DE  
XX Recombinant expression vector; transcription regulatory element;  
KW Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.  
XX  
XX Klebsiella pneumoniae.  
OS  
XX US6610836-B1.  
XX  
XX 26-AUG-2003.  
PD  
XX 27-JAN-2000; 2000US-00489039.  
KW

XX 29-JAN-1999; 99US-0117747P.  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX Breton GL, Osborne M;  
XX WPI; 2003-895346/82.  
XX P-PSDB; ABO64079.  
XX  
XX New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for  
XX preparing a vaccine composition against Klebsiella pneumoniae.  
XX  
XX Disclosure; SEQ ID NO 3425; 932pp; English.  
XX  
XX The invention describes a new isolated nucleic acid encoding a Klebsiella  
XX pneumoniae polypeptide. Also described are: a recombinant expression  
XX vector comprising the nucleic acid, operably linked to a transcription  
XX regulatory element; and a cell comprising the recombinant expression  
XX vector. The nucleic acid is useful for preparing a vaccine composition  
XX against Klebsiella pneumoniae. This sequence encodes a Klebsiella  
XX pneumoniae polypeptide of the invention  
XX  
XX Sequence 843 BP; 166 A; 274 C; 253 G; 150 T; 0 U; 0 Other;  
XX  
XX Query Match 72.5%; Score 17.4; DB 11; Length 843;  
XX Best Local Similarity 71.4%; Pred. No. 1.9e+03;  
XX Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX 4 CTGCTGCGCCGTCGACGCGSAC 24  
XX |||:|||||:|||||:  
XX 337 CTGCTGCGCCGTCGACGCGSAC 357  
XX  
XX RESULT 28  
XX ABD01715  
XX ID ABD01715 standard; DNA; 1026 BP.  
XX  
XX ABD01715;  
XX  
XX 29-JUL-2004 (first entry)  
XX  
XX Pseudomonas aeruginosa polynucleotide #319.  
XX  
XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
XX antibacterial.  
XX  
XX Pseudomonas aeruginosa.  
XX  
XX US6551795-B1.  
XX  
XX 22-APR-2003.  
XX  
XX 18-FEB-1999; 99US-002522991.  
XX  
XX 18-FEB-1998; 98US-0074788P.  
XX  
XX 27-JUL-1998; 98US-0094190P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX  
XX Rubenfeld MJ, Nolling J, Deloughery C, Bush D;  
XX  
XX WPI; 2003-615309/58.  
XX  
XX P-PSDB; ABO68144.  
XX  
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
XX useful as molecular targets for diagnostics, prophylaxis and treatment of  
XX pathological conditions resulting from bacterial infection.  
XX  
XX Disclosure; SEQ ID NO 319; 455pp; English.  
XX  
XX The invention relates to Pseudomonas aeruginosa polypeptides and the  
XX polynucleotides encoding them. The sequences are useful in diagnosis and

CC therapy of pathological conditions, as molecular targets for diagnostics,  
CC prophylaxis and treatment of pathological conditions resulting from a  
CC bacterial infection, for evaluating a compound, such as a polypeptide,  
CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
CC effective antibacterial targets, as targets for antibacterial drugs,  
CC including anti-P. aeruginosa drugs, as templates for recombinant  
CC production of P. aeruginosa-derived peptides or polypeptides, as target  
CC components for diagnosis and/or treatment of P. aeruginosa-caused  
CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biochip technology. Sequences ABD01397-  
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC seqdata.uspto.gov/sequence.html  
XX  
XX Sequence 1026 BP; 148 A; 400 C; 330 G; 147 T; 0 U; 1 Other;  
XX  
XX Query Match 72.5%; Score 17.4; DB 11; Length 1026;  
XX Best Local Similarity 71.4%; Pred. No. 1.9e+03;  
XX Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
XX  
XX 2 TCCTGTCGCGCCGTCGACGCGS 22  
XX |||:|||||:|||||:  
XX 78 TCCTGTCGCGCCGTCGACGCGS 98  
XX  
XX RESULT 29  
XX ABQ90070  
XX ID ABQ90070 standard; DNA; 1212 BP.  
XX  
XX ABQ90070;  
XX  
XX 01-OCT-2002 (first entry)  
XX  
XX M. capsulatus gene #55 for DNA array.  
XX  
XX Micro array; gene; ds; differential expression; gene expression.  
XX  
XX Methylococcus capsulatus.  
XX  
XX WO200255655-A2.  
XX  
XX 18-JUL-2002.  
XX  
XX 14-JAN-2002; 2002WO-N0000019.  
XX  
XX 12-JAN-2001; 2001NO-00000235.  
XX  
XX 12-JAN-2001; 2001NO-00000239.  
XX  
XX (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.  
XX (TIGR-) TIGR.  
XX  
XX Birkeland NK, Bidhammer I, Jonassen I, Jensen HB, Lien T,  
XX Lillehaug JR, Lossius I, Eisen JA, Frazer CM, Durkin AS,  
XX Salzberg SL;  
XX  
XX WPI; 2002-557818/59.  
XX  
XX Novel DNA array useful for determining differential expression of  
XX Methylococcus capsulatus genes, comprises polynucleotides or  
XX oligonucleotides representative for a selective number of Methylococcus  
XX capsulatus genes.  
XX  
XX Claim 19; Page 76-77; 678pp; English.  
XX  
XX The invention relates to a novel DNA array giving a representation of a  
XX number of Methylococcus capsulatus genes. The method of the invention is  
XX useful for determination of the differential expression of the genes of  
XX M. capsulatus, and for studying gene expression on a genomic scale and in  
XX gene expression assays of M. capsulatus genes. The sequences shown in  
XX ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the  
XX invention

SO Sequence 1212 BP; 233 A; 375 C; 384 G; 219 T; 0 U; 1 Other;

Query Match 72.5%; Score 17.4; DB 6; Length 1212;  
 Best Local Similarity 71.4%; Pred. No. 1.9e+03;  
 Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 4 CTGCTGCGCGGTCTGCGGAC 24  
 |||:|||||:|||||:  
 DB 823 CTGCTGCGCGGTCTGCGGAC 843

RESULT 30

ABD07361  
 ID ABD07361 standard; DNA; 1578 BP.

AC ABD07361;

XX 29-JUL-2004 (first entry)

DE Pseudomonas aeruginosa polynucleotide #5965.

XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
 antibacterial.

KW Pseudomonas aeruginosa.

OS US6551795-B1.

XX 22-APR-2003.

PD 18-FEB-1999; 99US-00252991.

XX 18-FEB-1998; 98US-0074788P.

PR 27-JUL-1998; 98US-0094190P.

XX (GENO-) GENOME THERAPEUTICS CORP.

PA Rubenfield MJ, Nolling J, Deloughery C, Bush D;

XX WPI: 2003-615309/58.

DR P-PSDB; ABO73790.

PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
 useful as molecular targets for diagnostics, prophylaxis and treatment of  
 PT pathological conditions resulting from bacterial infection.

XX Disclosure; SEQ ID NO 5965; 455bp; English.

CC The invention relates to Pseudomonas aeruginosa polypeptides and the  
 CC polynucleotides encoding them. The sequences are useful in diagnosis and  
 CC therapy of pathological conditions, as molecular targets for diagnostics,  
 CC prophylaxis and treatment of pathological conditions resulting from a  
 CC bacterial infection, for evaluating a compound, such as a polypeptide,  
 CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
 CC effective antibacterial targets, as targets for antibacterial drugs,  
 CC including anti-P. aeruginosa drugs, as templates for recombinant  
 CC production of P. aeruginosa-derived peptides or polypeptides, as target  
 CC components for diagnosis and/or treatment of P. aeruginosa-caused  
 CC infection, and in detection of P. aeruginosa sequences or other sequences  
 CC of Pseudomonas species using biochip technology. Sequences ABD01397-  
 CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification but was obtained in electronic format from USPTO at  
 CC seqdata.uspto.gov/sequence.html

XX Sequence 1578 BP; 192 A; 537 C; 530 G; 319 T; 0 U; 0 Other;

Query Match 72.5%; Score 17.4; DB 11; Length 1578;  
 Best Local Similarity 71.4%; Pred. No. 1.9e+03;  
 Matches 15; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 2 TCCTGCTGCGCGGTCTGCGGAC 22  
 |||:|||||:|||||:

DB 434 TCCTGCTGCGCGGTCTGCGGAC 454

Search completed: July 20, 2005, 16:11:57  
 Job time : 420.857 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2005, 14:42:10 ; Search time 1136.57 Seconds  
(without alignments)  
1023.187 Million cell updates/sec

Title: US-10-017-471B-2

Perfect score: 24  
Sequence: 1 gtcctgtgcgcgtacacgcgac 24

Scoring table: IDENTITY NUC  
Gapop 10\_0 , Gapext 1\_0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 300 summaries

Database :

GenEmbl1:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_stg:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
C 1	22	91.7	1215	1 AB001608 Streptomy
C 2	21	87.5	6252	1 SVU24659 Streptomy
C 3	20.4	85.0	2234	1 AB001683 Streptomy
C 4	20	83.3	93902	8 AP003767 Oryza sat
C 5	20	83.3	234050	1 AL627282 Salmoneil
C 6	20	81.7	1203	1 STMAFSAA M4250 S. griseus a
C 7	19.6	81.7	12070	1 AB011413 Streptomy
C 8	19.6	81.7	12070	1 AB011413 Streptomy
C 9	19.2	80.0	6727	1 SC0007731 Streptomy
C 10	19.2	80.0	6727	1 SC0007731 Streptomy
C 11	19.2	80.0	131711	2 AP003954 Oryza sat
C 12	19.2	80.0	145772	8 AP004269 Oryza sat
C 13	19.2	80.0	230850	1 SC0393127 Streptomy
C 14	18.8	78.3	1851	10 BC032284 Mus muscu
C 15	18.8	78.3	1851	10 BC032284 Mus muscu
C 16	18.8	78.3	1901	10 AB036748 Mus muscu
C 17	18.8	78.3	1904	10 AB036746 Mus muscu
C 18	18.8	78.3	1919	10 AB036749 Mus muscu
C 19	18.8	78.3	8992	1 SV117268 Streptomy

C 20	18.8	78.3	11004	1 AE012334 Xanthomon
C 21	18.8	78.3	13513	1 AF506520 Streptomy
C 22	18.8	78.3	15356	1 AE004867 Pseudomon
C 23	18.8	78.3	34398	1 AY258138 Pseudomon
C 24	18.8	78.3	50550	7 AF271693 Mycobacte
C 25	18.8	78.3	94058	2 AC141024 Rattus no
C 26	18.8	78.3	110000	1 AP006618_13
C 27	18.8	78.3	146585	8 CDS0808CAM
C 28	18.8	78.3	152186	2 AC084284
C 29	18.8	78.3	162995	10 AL663032
C 30	18.8	78.3	193050	2 AL646062
C 31	18.8	78.3	199271	2 AC142243 Mus muscu
C 32	18.8	78.3	229737	2 AC097884
C 33	18.8	78.3	303855	1 AE017230
C 34	18.4	76.7	1850	1 TSU17342
C 35	18.4	76.7	3168	6 AX654752
C 36	18.4	76.7	5367	1 SCABACG2
C 37	18.4	76.7	9090	1 AY392413
C 38	18.4	76.7	9316	1 AF077869
C 39	18.4	76.7	11069	1 AE008208
C 40	18.4	76.7	11293	1 AE009406
C 41	18.4	76.7	12358	1 AE012437 Xanthomon
C 42	18.4	76.7	24720	1 AJ632203 Streptomy
C 43	18.4	76.7	110000	1 AP006618_18
C 44	18.4	76.7	110000	1 AP006618_39
C 45	18.4	76.7	139773	2 AC067806
C 46	18.4	76.7	140226	2 AC067806
C 47	18.4	76.7	343473	9 AE012461 Xanthomon
C 48	18.2	75.8	84248	1 AP001783
C 49	18.2	75.8	140140	2 AC151647
C 50	18.2	75.8	153699	2 AC144285
C 51	18.2	75.8	165005	2 AC009940
C 52	18.2	75.8	250792	2 AC096277 Rattus no
C 53	18	75.0	25306	1 AF058302
C 54	18	75.0	110000	2 AP006490_2
C 55	17.8	74.2	309	12 AIG57179
C 56	17.8	74.2	1455	3 AK116370
C 57	17.8	74.2	1521	1 AMU3061
C 58	17.8	74.2	3573	1 AB016763 Streptomy
C 59	17.8	74.2	11057	1 AE012461 Xanthomon
C 60	17.8	74.2	12633	1 AE013939
C 61	17.8	74.2	16268	1 AE004753
C 62	17.8	74.2	21562	1 PPS79894
C 63	17.8	74.2	51975	9 AC108475
C 64	17.8	74.2	98734	9 AC013277
C 65	17.8	74.2	109528	1 AF040570
C 66	17.8	74.2	110000	1 AE017282_09
C 67	17.8	74.2	110000	1 AE017282_13
C 68	17.8	74.2	110000	1 AP006618_25
C 69	17.8	74.2	110000	1 BX571965_14
C 70	17.8	74.2	110000	1 BX936398_38
C 71	17.8	74.2	110000	8 CR382129_07
C 72	17.8	74.2	205544	2 CR847834
C 73	17.8	74.2	208050	1 AJ414145
C 74	17.8	74.2	210614	1 AB088224 Streptomy
C 75	17.8	74.2	226349	2 AC140865
C 76	17.8	74.2	238543	2 CR456635
C 77	17.8	74.2	280558	1 AE017301
C 78	17.8	74.2	289555	1 AP005947
C 79	17.8	74.2	290333	1 AE017140 Yersinia
C 80	17.8	74.2	299800	1 AP005028 Streptomy
C 81	17.8	74.2	310174	1 AE016870 Pseudomon
C 82	17.6	73.3	452	1 AF010105
C 83	17.6	73.3	460	1 AY651310
C 84	17.6	73.3	498	1 AY214755
C 85	17.6	73.3	1419	1 AF134587 Streptomy
C 86	17.6	73.3	1482	1 AB117717 Unculture
C 87	17.6	73.3	9971	1 AE000820 Methanoba
C 88	17.6	73.3	45503	1 AY196994 Streptomy
C 89	17.6	73.3	105841	6 AC092697 Oryza sat
C 90	17.6	73.3	110000	1 BX571966_08
C 91	17.6	73.3	110000	1 CP000011_08
C 92	17.6	73.3	162215	9 AC004896 Homo sapi

C 93	17.6	73.3	194361	8	AC092289	AC092289	Oryza sat	C 166	17.2	71.7	4511	6	CO729197	CO729197	Sequence
C 94	17.6	73.3	201201	8	AC150433	AC150433	Colobus g	C 167	17.2	71.7	4511	9	AB023202	AB023202	Homo sapi
C 95	17.6	73.3	300029	2	AB017115	AB017115	Oryza sat	C 168	17.2	71.7	5137	9	AK090454	AK090454	Homo sapi
C 96	17.6	73.3	300100	1	SC0939123	SC0939123	Streptococ	C 169	17.2	71.7	5344	9	HSN607425	HSN607425	Homo sapi
C 97	17.6	73.3	302070	1	AP005823	AP005823	Corynebact	C 170	17.2	71.7	5373	3	SADNMTW5	SADNMTW5	Streptococ
C 98	17.6	73.3	310550	1	SC0939113	SC0939113	Streptococ	C 171	17.2	71.7	6523	1	AB017045	AB017045	Paracoccu
C 99	17.4	72.5	843	6	AR386696	AR386696	Sequence	C 172	17.2	71.7	6653	1	AY162971	AY162971	Microsom
C 100	17.4	72.5	1212	6	AR363420	AR363420	Sequence	C 173	17.2	71.7	7055	1	ASR457162	ASR457162	Arthrobac
C 101	17.4	72.5	1688	1	STMSUBCB	STMSUBCB	Sequence	C 174	17.2	71.7	7403	8	AF047459	AF047459	Chlamydia
C 102	17.4	72.5	1998	6	AR363414	AR363414	Sequence	C 175	17.2	71.7	7860	1	SKZ86111	SKZ86111	Streptococ
C 103	17.4	72.5	2351	6	CO842135	CO842135	Sequence	C 176	17.2	71.7	7980	1	MXU20669	MXU20669	Myxococcu
C 104	17.4	72.5	2351	9	AK125166	AK125166	Sequence	C 177	17.2	71.7	8568	8	AF449619	AF449619	Phytophth
C 105	17.4	72.5	9828	1	AB004468	AB004468	Pseudomon	C 178	17.2	71.7	10136	1	AF546156	AF546156	Microsom
C 106	17.4	72.5	10202	1	AB004483	AB004483	Pseudomon	C 179	17.2	71.7	10932	1	BCPYR	BCPYR	B. caldolyti
C 107	17.4	72.5	11842	1	AB005053	AB005053	Halobacte	C 180	17.2	71.7	11092	1	AE005703	AE005703	Caillodact
C 108	17.4	72.5	11886	1	AB004638	AB004638	Pseudomon	C 181	17.2	71.7	12675	1	AE001878	AE001878	Deinococc
C 109	17.4	72.5	47344	9	AF061779	AF061779	Homo sapi	C 182	17.2	71.7	12908	1	AE004934	AE004934	Pseudomon
C 110	17.4	72.5	97556	9	AC079781	AC079781	Homo sapi	C 183	17.2	71.7	12995	1	AE001857	AE001857	Xanthomon
C 111	17.4	72.5	99549	2	AC092862_3	AC092862_3	Continuatio	C 184	17.2	71.7	13056	1	AE004713	AE004713	Pseudomon
C 112	17.4	72.5	110000	2	AE017282_02	AE017282_02	Continuatio	C 185	17.2	71.7	13268	6	AX587577	AX587577	Sequence
C 113	17.4	72.5	110000	2	AC119727_1	AC119727_1	Continuatio	C 186	17.2	71.7	13528	1	AE014340	AE014340	Brucella
C 114	17.4	72.5	110121	9	AL590989	AL590989	Human DNA	C 187	17.2	71.7	13556	1	AE009600	AE009600	Brucella
C 115	17.4	72.5	117296	9	AC092865	AC092865	Homo sapi	C 188	17.2	71.7	13857	6	AR338584	AR338584	Streptococ
C 116	17.4	72.5	135431	9	AC106847	AC106847	Homo sapi	C 189	17.2	71.7	14219	1	AB070957	AB070957	Streptococ
C 117	17.4	72.5	143717	2	AC083933	AC083933	Homo sapi	C 190	17.2	71.7	14467	9	CVU84760	CVU84760	Chromatium
C 118	17.4	72.5	166246	2	AF271408	AF271408	Homo sapi	C 191	17.2	71.7	17019	9	EX936374	EX936374	Human DNA
C 119	17.4	72.5	167576	2	AC147297	AC147297	Pan trogl	C 192	17.2	71.7	18977	1	SAR7932	SAR7932	Streptococ
C 120	17.4	72.5	167875	2	AC083906	AC083906	Homo sapi	C 193	17.2	71.7	27521	1	AB112586	AB112586	Streptococ
C 121	17.4	72.5	167914	2	AC055870	AC055870	Homo sapi	C 194	17.2	71.7	30000	6	AX250263	AX250263	Sequence
C 122	17.4	72.5	167675	2	AC016978	AC016978	Homo sapi	C 195	17.2	71.7	34869	1	AF324838	AF324838	Streptococ
C 123	17.4	72.5	185636	2	AC145875	AC145875	Pan trogl	C 196	17.2	71.7	36602	6	CO878999	CO878999	Sequence
C 124	17.4	72.5	281450	1	AP005032	AP005032	Streptococ	C 197	17.2	71.7	36748	7	AY539836	AY539836	Burkholder
C 125	17.4	72.5	303206	2	AC122945	AC122945	Rattus no	C 198	17.2	71.7	38846	2	AC145717	AC145717	Homo sapi
C 126	17.4	72.5	332889	2	AC098450	AC098450	Rattus no	C 199	17.2	71.7	39568	2	AC014823	AC014823	Drosophila
C 127	17.4	72.5	339650	1	SC0939108	SC0939108	Streptococ	C 200	17.2	71.7	39753	2	AC149318	AC149318	PhaKoposor
C 128	17.2	71.7	318	9	HSB17868	HSB17868	Homo sapien	C 201	17.2	71.7	46777	9	EX247883	EX247883	Human DNA
C 129	17.2	71.7	453	12	AY657216	AY657216	Synthetic	C 202	17.2	71.7	51915	9	AP005273	AP005273	Homo sapi
C 130	17.2	71.7	498	9	AF424846	AF424846	Macaca mu	C 203	17.2	71.7	57339	9	AL928593	AL928593	Human DNA
C 131	17.2	71.7	603	6	AX196096	AX196096	Sequence	C 204	17.2	71.7	82746	1	AF453501	AF453501	Actinobyn
C 132	17.2	71.7	708	8	AU066535	AU066535	Chlamydom	C 205	17.2	71.7	88421	6	AX417445	AX417445	Sequence
C 133	17.2	71.7	825	6	CO594282	CO594282	Sequence	C 206	17.2	71.7	96153	2	CR755972	CR755972	Homo sapi
C 134	17.2	71.7	861	3	AY118988	AY118988	Drosophila	C 207	17.2	71.7	96335	9	EX679671	EX679671	Human sapi
C 135	17.2	71.7	1415	1	THR295159	THR295159	Thermus t	C 208	17.2	71.7	100267	9	HSNHC3W6A	HSNHC3W6A	Homo sapien
C 136	17.2	71.7	1500	5	BC065586	BC065586	Danio rer	C 209	17.2	71.7	102750	9	AC004465	AC004465	Homo sapi
C 137	17.2	71.7	1632	6	AX654521	AX654521	Sequence	C 210	17.2	71.7	103450	1	AF440781	AF440781	Streptococ
C 138	17.2	71.7	1826	1	ABN1FH	ABN1FH	Sequence	C 211	17.2	71.7	104677	9	CR753803	CR753803	Human DNA
C 139	17.2	71.7	1934	9	BC009843	BC009843	Homo sapi	C 212	17.2	71.7	106117	2	AP000628	AP000628	Homo sapi
C 140	17.2	71.7	1968	9	BC002605	BC002605	Homo sapi	C 213	17.2	71.7	107413	4	AL773560	AL773560	Pig DNA
C 141	17.2	71.7	2047	8	AK110881	AK110881	Oryza sat	C 214	17.2	71.7	109519	6	AX195929	AX195929	Sequence
C 142	17.2	71.7	2070	6	CO879003	CO879003	Sequence	C 215	17.2	71.7	109528	1	AF040570	AF040570	Amvcolato
C 143	17.2	71.7	2127	6	CO586451	CO586451	Sequence	C 216	17.2	71.7	109646	9	HSNHC78S22	HSNHC78S22	Homo sapi
C 144	17.2	71.7	2189	6	AX834331	AX834331	Sequence	C 217	17.2	71.7	110000	1	AE016822_06	AE016822_06	Continuatio
C 145	17.2	71.7	2189	9	AK096721	AK096721	Homo sapi	C 218	17.2	71.7	110000	1	AE017282_00	AE017282_00	Methyloco
C 146	17.2	71.7	2454	8	AF022816	AF022816	Chlamydom	C 219	17.2	71.7	110000	1	AE017282_13	AE017282_13	Continuatio
C 147	17.2	71.7	2648	1	SCSPRAGE	SCSPRAGE	S. coelicolo	C 220	17.2	71.7	110000	1	AE017282_16	AE017282_16	Continuatio
C 148	17.2	71.7	2688	6	CO879005	CO879005	Sequence	C 221	17.2	71.7	110000	1	AP006618_17	AP006618_17	Continuatio
C 149	17.2	71.7	2712	6	PDEPHAAC	PDEPHAAC	Paracoccu	C 222	17.2	71.7	110000	1	AP006618_23	AP006618_23	Continuatio
C 150	17.2	71.7	2825	6	CO594282	CO594282	Sequence	C 223	17.2	71.7	110000	1	AP006618_43	AP006618_43	Continuatio
C 151	17.2	71.7	2825	6	AB015139	AB015139	Chlamydom	C 224	17.2	71.7	110000	1	AP006618_44	AP006618_44	Continuatio
C 152	17.2	71.7	2867	8	AB015139	AB015139	Chlamydom	C 225	17.2	71.7	110000	1	AP006618_45	AP006618_45	Continuatio
C 153	17.2	71.7	2888	6	E34218	E34218	Chlorophyll	C 226	17.2	71.7	110000	1	AP006618_52	AP006618_52	Continuatio
C 154	17.2	71.7	2963	9	BC017259	BC017259	Homo sapi	C 227	17.2	71.7	110000	1	AP006840_01	AP006840_01	Continuatio
C 155	17.2	71.7	3042	1	AB091059	AB091059	Glucanase	C 228	17.2	71.7	110000	1	AP006840_17	AP006840_17	Continuatio
C 156	17.2	71.7	3146	1	AY348315	AY348315	Streptococ	C 229	17.2	71.7	110000	1	AP006840_22	AP006840_22	Continuatio
C 157	17.2	71.7	3271	3	SD0580406	SD0580406	Subterfices	C 230	17.2	71.7	110000	2	AP006496_3	AP006496_3	Continuatio
C 158	17.2	71.7	3292	1	BSA17805	BSA17805	Bacillus	C 231	17.2	71.7	115292	9	AC026425	AC026425	Homo sapi
C 159	17.2	71.7	3542	1	AB071166	AB071166	Glucanase	C 232	17.2	71.7	115292	2	AC120117	AC120117	Homo sapi
C 160	17.2	71.7	3835	9	HSNHC71B	HSNHC71B	Human MHC	C 233	17.2	71.7	119151	2	AP000655	AP000655	Homo sapi
C 161	17.2	71.7	4050	1	STW1LVBNC	STW1LVBNC	Streptococ	C 234	17.2	71.7	120628	9	AC012153	AC012153	Human sapi
C 162	17.2	71.7	4133	1	AB054887	AB054887	Streptococ	C 235	17.2	71.7	129811	9	HSJ344F7	HSJ344F7	Human DNA
C 163	17.2	71.7	4142	9	AB124552	AB124552	Homo sapi	C 236	17.2	71.7	130141	9	AL513320	AL513320	Human DNA
C 164	17.2	71.7	4201	1	AY588479	AY588479	Pseudomon	C 237	17.2	71.7	135638	9	AF484556	AF484556	Streptococ
C 165	17.2	71.7	4210	1	MSG1VB	MSG1VB	Mycobacteri	C 238	17.2	71.7	138992	9	AL662828	AL662828	Human DNA

239	17.2	71.7	142560	2	CR753845
C 240	17.2	71.7	145299	2	AF228728
C 241	17.2	71.7	146299	2	AL359826
C 242	17.2	71.7	146436	2	AC108759
C 243	17.2	71.7	147891	8	AP004693
C 244	17.2	71.7	149303	8	AC087412
C 245	17.2	71.7	150925	9	AC113607
C 246	17.2	71.7	150925	9	AC125282
C 247	17.2	71.7	155503	2	AC084347
C 248	17.2	71.7	155159	9	AL591916
C 249	17.2	71.7	160489	5	BX005126
C 250	17.2	71.7	162785	3	BX547938
C 251	17.2	71.7	163279	3	AC104508
C 252	17.2	71.7	165079	2	AC024517
C 253	17.2	71.7	167300	2	AC021394
C 254	17.2	71.7	169259	2	AP006288
C 255	17.2	71.7	169740	2	AF186193
C 256	17.2	71.7	169856	2	AC144890
C 257	17.2	71.7	170859	10	AC128952
C 258	17.2	71.7	172249	2	CR753839
C 259	17.2	71.7	176267	4	AC148660
C 260	17.2	71.7	179691	4	BX296515
C 261	17.2	71.7	180559	9	AL645922
C 262	17.2	71.7	181210	9	AC011742
C 263	17.2	71.7	181858	9	AC084856
C 264	17.2	71.7	182705	3	AC091208
C 265	17.2	71.7	184418	2	AC130644
C 266	17.2	71.7	186024	2	AC084225
C 267	17.2	71.7	186335	2	BX927323
C 268	17.2	71.7	187081	10	AC124421
C 269	17.2	71.7	188700	2	AC021603
C 270	17.2	71.7	189050	1	AL646066
C 271	17.2	71.7	189050	1	AL646069
C 272	17.2	71.7	189424	9	AC148664
C 273	17.2	71.7	193915	2	AC091457
C 274	17.2	71.7	199204	2	AC131298
C 275	17.2	71.7	203613	10	AL591404
C 276	17.2	71.7	207206	9	AC090260
C 277	17.2	71.7	210018	2	AC015557
C 278	17.2	71.7	213050	1	AL646079
C 279	17.2	71.7	213201	2	BX927095
C 280	17.2	71.7	215210	2	CR749744
C 281	17.2	71.7	218445	2	AC083959
C 282	17.2	71.7	219682	2	AC150692
C 283	17.2	71.7	224640	2	AC125590
C 284	17.2	71.7	250762	2	AL662840
C 285	17.2	71.7	257843	2	AC095980
C 286	17.2	71.7	272233	2	AC106289
C 287	17.2	71.7	273285	1	AE017304
C 288	17.2	71.7	273769	2	AC099070
C 289	17.2	71.7	295150	1	SC0939125
C 290	17.2	71.7	298300	1	AP005025
C 291	17.2	71.7	299050	1	SC0939119
C 292	17.2	71.7	299300	1	AP005026
C 293	17.2	71.7	299986	1	AE017240
C 294	17.2	71.7	300511	1	AE016775
C 295	17.2	71.7	300861	1	AE016777
C 296	17.2	71.7	301399	1	AE017233
C 297	17.2	71.7	302007	1	SC0939132
C 298	17.2	71.7	302325	1	AE017236
C 299	17.2	71.7	302325	1	AE017236
C 300	17.2	71.7	302898	1	AE017238

## ALIGNMENTS

RESULT 1  
LOCUS AB001608/c 1215 bp DNA linear BCT 22-NOV-1997  
DEFINITION Streptomyces virginiae DNA for BarX, complete cds.  
ACCESSION AB001608  
VERSION AB001608.1 GI:2641955

KEYWORDS BarX.  
SOURCE Streptomyces virginiae  
ORGANISM Streptomyces virginiae  
REFERENCE 1 (ntes)  
AUTHORS Kinoshita, H., Iyoshi, H., Okamoto, S., Nakano, H., Nihira, T. and Yamada, Y.  
TITLE Butyrolactone autoregulator receptor protein (BarA) as a transcriptional regulator in Streptomyces virginiae  
JOURNAL J. Bacteriol. 179 (122), 6986-6993 (1997)  
MEDLINE 98037495  
PUBMED 9371444  
REFERENCE 2 (bases 1 to 1215)  
AUTHORS Kinoshita, H.  
TITLE Direct Submission  
JOURNAL Submitted (06-MAR-1997) Hiroshi Kinoshita, Osaka University, Department of Engineering, Yamadaoka 2-1, Suita, Osaka 565, Japan (E-mail: kinoshita@biochem.bio.eng.osaka-u.ac.jp, Tel: +81-6-879-7433, Fax: +81-6-879-7432)  
FEATURES  
source location/Qualifiers  
1. .1215  
/organism="Streptomyces virginiae"  
/mol\_type="genomic DNA"  
/DB\_xref="taxon:1961"  
244. .1128  
/gene="barX"  
244. .1128  
/gene="barX"  
/codon\_start=1  
/transl\_table=1  
/product="BarX"  
/protein\_id="BAA23611.1"  
/db\_xref="GI:2641956"  
/translation="MTSTVPRILVHRAVAVEFLVQMSPTAENRPAALTQMPRAHYF  
TPVNGCDPILASBITRQVGTLLSHAFRGVSFQDQLMDLHNSVPRQAGVAAAPD  
LELDVICSDFIRRRGRRLAGRYEVTLVYCGQVATATGAADFCTSPVYTORLGRDVG  
TGVRPLPOPLAPSVGRFLTVDVLSATERPLEMQRVDEQHPVLDHPDHPGVGL  
MESARPLAOAIDPSRPFLLPTMRSEFSRYAELDRPCMIQNEPLPADNGDRQVRVVGH  
QDDTVFSCILGRGAAB"

## ORIGIN

Query Match 91.7%; Score 22; DB 1; Length 1215;  
Best Local Similarity 79.2%; Pred. No. 4.4e+02;  
Matches 19; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTCTGTCGCGCGTACGCGSAC 24  
Db 1077 GTCTGTGTGCGCGTACGCGGAC 1054

## RESULT 2

SVU24659/c 6252 bp DNA linear BCT 07-AUG-2003  
LOCUS SVU24659  
DEFINITION Streptomyces venezuelae jad gene cluster, complete sequence.  
ACCESSION U24659  
VERSION U24659.2 GI:16445340  
KEYWORDS Streptomyces venezuelae  
SOURCE Streptomyces venezuelae  
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomyces; Streptomyces; Streptomyces; Streptomyces.  
REFERENCE 1 (bases 2899 to 6252)  
AUTHORS Yang, K., Han, L., and Vining, L.C.  
TITLE Regulation of jadomycin B production in Streptomyces venezuelae  
JOURNAL J. Bacteriol. 177 (21), 6111-6117 (1995)  
MEDLINE 96042086  
PUBMED 7592375  
REFERENCE 2 (bases 2899 to 6252)  
AUTHORS Yang, K., Han, L., Vining, L.C., and He, J.Y.  
TITLE Participation of jadR in the regulation of jadomycin B production in Streptomyces venezuelae ISP 5230

```

JOURNAL      Unpublished
REFERENCE    3 (bases 2899 to 6252)
AUTHORS      Yang, K.
TITLE        Direct Submission
JOURNAL      Submitted (11-APR-1995) Kegan Yang, Biology, Dalhousie University,
AUTHORS      Halifax, Nova Scotia B3H 4J1, Canada
REFERENCE    4 (bases 1 to 6252)
AUTHORS      Wang, L. and Vining, L.C.
TITLE        Control of growth, secondary metabolism and sporulation in
              Streptomyces venezuelae ISP5230 by jadW1), a member of the afsA
              family of gamma-butyrolactone regulatory genes
JOURNAL      Microbiology (Reading, Engl.) 143 (Pt 8), 1991-2004 (2003)
MEDLINE      22787062
PUBMED       12904539
REFERENCE    5 (bases 1 to 6252)
AUTHORS      Wang, L. and Vining, L.C.
TITLE        Direct Submission
JOURNAL      Submitted (26-OCT-2001) Biology, Dalhousie University, 1355 Oxford,
              Halifax, Nova Scotia B3H 4J1, Canada
REMARK       Sequence update by submitter
COMMENT      On Oct 26, 2001 this sequence version replaced gi:886035.
FEATURES     location/Qualifiers
              1..6252
              /organism="Streptomyces venezuelae"
              /mol_type="genomic DNA"
              /strain="ISP5230"
              /db_xref="taxon:54571"
              /note="wild type"
              195..1124
              /gene="jadW1"
              /note="similar to Barx and AfsA"
              195..1124
              /gene="jadW1"
              /codon_start=1
              /transl_table=11
              /product="JadW1"
              /protein_id="AAB36582.2"
              /db_xref="GI:16445341"
              /translation="MHTSRGEFVRADPADIIIPDWTOLRONRFSVSARVACLSL
              LSRAGARHPMLVAETIPETSMVAHAELGVLDQFVWMDLSYADSEALVLDGLS
              DVTYVVCSDITRRSGRLRNLTYYVLRDDLLATGSGTARCTALAYRRGGRME
              ALGRPEVPLIPGVPRIVGARTEDEVLAENKPDQMLRNTAHTTLFRRPDHYECM
              VILLEAARQAATATGSAAYLPDLSVSPFLRYVELDSPCMIEASVPTDPSTTTIRVT
              GHGDSPVFRCTLSPSRELSVATGALDTRLG"
              1139..2086
              /gene="jadW2"
              /gene="jadW2"
              1139..2086
              /gene="jadW2"
              /function="transcriptional regulation"
              /note="dehydrogenase; repressor"
              /codon_start=1
              /transl_table=11
              /product="JadW2"
              /protein_id="AAL23835.1"
              /db_xref="GI:16445342"
              /translation="MGPRLITGATGFGIGRVAAAAATAREPAHYRLLAGPLTPAA
              AAGPRETAVGDLRDPASLRACGEVDVLIHCASAIIGDAELARVNDGENTLADAA
              VNAAGRTVGPBHRILRTRGPPRAAPGTLIPASVSSTPAAAGRRVMDGCVIVR
              PHVVGTRRLPLGLVRLAALPEPLDGGSSLSHSLIVDSILAGLLGAAISPAIRPG
              PYYVDHPEVPVGDLLAAAESLLTRGLPRGAVALDPTARRLAGLPPAGHLLMLAT
              DMFADRPWRRELCPPGGLPRGSRTRRTGTTDSITP"
              2109..2873
              /note="dehydrogenase/reductases; regulator for jadomycin B
              biosynthesis"
              /codon_start=1
              /transl_table=11
              /product="JadW3"
              /protein_id="AAL23836.1"
              /db_xref="GI:16445343"
              /translation="MSALGRTRAVTGSARGIGRIARLADGALVAVHYSNKA
              LETVELIERGGARFAFRAELGAPDVAVDVTFYALDGLERGAAREFDILVNNAAAG
              SRIHQLTTEVDRFLPAIVNKAPLVVGRLDRDGRIVNISSAATGAPDSDVTY
              AATKGAVDMTTALAKELGPRGITVNAVAPGYIADDMNRRRTPEASAAALAMSVFN

```

```

gene          RIGTPADVGRRPRVSDERWITGVYDATTGATL"
              complement (2953..3543)
              /gene="jadR2"
              /complement (2953..3543)
              /gene="jadR2"
              /function="transcriptional regulation"
              /note="repressor"
              /citation=11
              /codon_start=1
              /evidence=experimental
              /transl_table=11
              /product="JadR2"
              /protein_id="AAB36583.1"
              /db_xref="GI:886037"
              /translation="MTQKQARTRTRALISAAEFDEHGVALAKLSAISSGACVSG
              ALHFHPENKVAAYEIDASTTLRTARIIVHOSNNAQNLADYTHALARLVREDVVR
              AGFLSCSOLCGTDNLNROEMSCVCGORLAEDGGLADIGGQDLARTIVATIG
              LEALCRNGEMLSPGIVTGLMRTLPIVAAPGRSP"
              4192..4896
              /gene="jadR1"
              4192..4896
              /gene="jadR1"
              /function="transcriptional regulation"
              /note="response regulator"
              /codon_start=1
              /evidence=experimental
              /transl_table=11
              /product="JadR1"
              /protein_id="AAB36584.1"
              /db_xref="GI:886038"
              /translation="MTVDHIGEGGPTTAAGRVLYVESCGKEGGLVQGRRHGHV
              DRVETGGALQAYEADLVLIDELPLDGLVEVGRIGRSADVPVIAVTSKSLDXY
              LGLOAGADYLVKPEYFRELMARMEAWRRARPASPAKRTTRGPHIDAAARCVTD
              GQVDLFRKEFDLXYLASHPDVTPKQOLMOQWGDWSRRRTVDTVSSLRNLTGAS
              DWVTITRGVFRFRG"

ORIGIN
Query Match      87.5%; Score 21; DB 1; Length 6252;
Best Local Similarity 78.3%; Pred. No. 8.2e+02;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY      1 GTCCGTGTCGCCGCGTSGACSCGSA 23
        |||||:|||||:|||||:|||||:|
Db      1034 GTCCGTGTCGCCGCGTSGACSCGSA 1012

RESULT 3
LOCUS      AB001683 2234 bp DNA linear BCT 22-AUG-1997
DEFINITION Streptomyces sp. gene for Farx, Farx, complete cds.
ACCESSION  AB001683
VERSION    AB001683.1 GI:2342428
KEYWORDS   Farx; Farx.
SOURCE     Streptomyces sp.
            Streptomyces sp.
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE  1 (sites)
AUTHORS    Waki, M., Mihara, T. and Yamada, Y.
TITLE      Cloning and characterization of the gene (farx) encoding the
            receptor for an extracellular regulatory factor (IM-2) from
            Streptomyces sp. strain FRI-5
JOURNAL    J. Bacteriol. 179 (16), 5131-5137 (1997)
MEDLINE    97405912
PUBMED     9260956
REFERENCE  2 (bases 1 to 2234)
AUTHORS    Waki, M.
TITLE      Direct Submission
JOURNAL    Submitted (07-MAR-1997) Miyoko Waki, Osaka University, Graduate
            School of Engineering, Department of Biotechnology, Yamadaoka 2-1,
            Suita, Osaka 565, Japan (E-mail:waki.yam@stu.bio.eng.osaka-u.ac.jp,
            Tel:+81-6-879-7433, Fax:+81-6-879-7432)
FEATURES   location/Qualifiers

```



source 1. .2234  
/organism="Streptomyces sp. "  
/mol\_type="genomic DNA"  
/strain="FRI-5"  
/db\_xref="taxon:1931"  
440. .1315  
/gene="farx"  
440. .1315  
/gene="farx"  
/codon\_start=1  
/transl\_except=(pos:440. .442, aa:Met)  
/transl\_table=1  
/product="Parx"  
/protein\_id="BA21185.1"  
/db\_xref="GI:2342429"  
/translation="MWRSTSAQVLLTDMORLDARFSVTARWPLSHAFPTVPGDGY  
DPLMCASTIRQIAYLVHSHAEVAFPGHQFVLMDSVSVRELLRVLVPAVDLATT  
CVETKRRARGLSGYEAVRRCQVAVTGAATCTSPAYGRIREHYLTPHRPL  
PLTPKRAPOSVARSPLTDVVLSPDLRENRQLRVDTNHPVLFDHWVDHVGMYLMEAA  
RQAAASALGRSPRPLGVAGFKRYVELDAECVIESRLFDVYGAEEVRAVYTGONG  
ELTVGVVTAASYG"  
1500. .2165  
/gene="fara"  
1500. .2165  
/gene="fara"  
/codon\_start=1  
/transl\_except=(pos:1500. .1502, aa:Met)  
/transl\_table=1  
/product="Para"  
/protein\_id="BA21185.1"  
/db\_xref="GI:2342430"  
/translation="MAEQVRAIRTRQALILSAANVPDEBGQAATTSIBLTIVAGTYG  
ALYHAFQSKEDLAQGVLTAAQEDLLPERPAKQEVVDVWLHTHRLRTPMVAGAR  
LSLVNAGAGLDRSAPFRNWDKFTDLLEKAQAGELLPHVVPATADVITGAYGVOS  
MSQALTEHQDGRVNVALLRHMPSIAQPSVLASLHIGESRAEEVYLEARQLAEQAD  
EED".

Query Match 85.0%; Score 20.4; DB 1; Length 2234;  
Best Local Similarity 75.0%; Pred. No. 1.6e+03;  
Matches 18; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GTCCTGSRGCGCGTSAACGCGSAC 24  
|||:||||:||||:||||:||||:||||:  
Db 1264 GTTCTGGTGGCGCGGTCAACGCGAC 1241

RESULT 4  
AP003767/c 93902 bp DNA linear PLN 23-APR-2004  
LOCUS Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 6,  
DEFINITION PAC clone: P0036F10, complete sequence.  
ACCESSION AP003767  
VERSION AP003767.2 GI:46518322  
SOURCE HTG.  
ORGANISM Oryza sativa (japonica cultivar-group)  
KEYWORDS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Euharidiales; Oryzaceae; Oryza.  
REFERENCE 1  
Saseaki, T., Matsumoto, T. and Yamamoto, K.  
Oryza sativa nipponbare (Gaz) genomic DNA, chromosome 6, PAC  
clone: P0036F10  
Published Only in Database (2001)  
2 (bases 1 to 93902)  
Saseaki, T., Matsumoto, T. and Yamamoto, K.  
Direct Submission  
Submitted (20-JUN-2001) Takuji Saseaki, National Institute of  
Agrobiological Sciences, Rice Genome Research Program, Kannondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail: tsaseaki@nias.affrc.go.jp, url: http://rpg.dna.affrc.go.jp/,  
tel: 81-298-7441, fax: 81-298-38-7468)

```

COMMENT
On Apr 22, 2004 this sequence version replaced gi:14517641.
The orientation of the sequence is from T7 to SP6 of the PAC clone.

FEATURES
    source
        1..93902
            /organism="Oryza sativa (japonica cultivar-group)"
            /mol_type="genomic DNA"
            /culivar="Nipponbare"
            /db_xref="taxon:39947"
            /chromosome="6"
            /clone="P0036F10"

ORIGIN
Query Match      83.3%; Score 20; DB 8; Length 93902;
Best Local Similarity 77.3%; Pred. No. 1,3e+03;
Matches 17; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY      3 CCTGCTGCGCCGCTGACGCGSAC 24
        |||:|||||:|||||:|||||:
Db      27007 CCTGTGTGCGCGTCAACCGGAC 26986

RESULT 5
AL627282/c AL627282 DNA linear BCT 04-JUL-2003
DEFINITION Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
LOCUS complete chromosome, segment 18/20.
ACCESSION AL627282 AL513382
VERSION AL627282.1 GI:16505159
KEYWORDS
SOURCE
ORGANISM
    Salmonella enterica subsp. enterica serovar Typhi
    Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
    Enterobacteriaceae; Salmonella.
    1 (bases 1 to 234050)
REFERENCE
    Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D.,
    Wain, J., Churcher, S., Mungall, K.L., Bentley, S.D., Holden, M.T.G.,
    Sebatia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T.,
    Connor, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N.,
    Farrar, J., Felwell, T., Hamlin, N., Haque, A., Hien, T.T., Holroyd, S.,
    Jagels, K., Krogh, A., Larsen, T.S., Leather, S., Moule, S., O'Gaora, P.,
    Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J.,
    Stevens, K., Whitehead, S. and Barrrell, B. G.
    Complete genome sequence of a multiple drug resistant Salmonella
    enterica serovar Typhi CT18
    Nature 413 (6858), 848-852 (2001)
JOURNAL
    21534947
REFERENCE
    11677608
    2 (bases 1 to 234050)
    Parkhill, J.
    Direct Submission
    Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
    sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
    Hinxton, Cambridge CB10 1SA, UK
    E-mail: parkhill@sanger.ac.uk
COMMENT
Notes:
Details of S. typhi sequencing at the Sanger Centre are available
on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S\_typhi/).
location/Qualifiers
    1..234050
        /organism="Salmonella enterica subsp. enterica serovar
        Typhi"
        /mol_type="genomic DNA"
        /strain="CT18"
        /db_xref="taxon:90370"
    92..463
        /gene="rplN"
        /note="synonym: STY4368"
    92..463
        /gene="rplN"
        /note="Orthologue of E. coli rplN (RL14 ECO1); Pasta hit
        to RL14 ECO1 (123 aa), 98% identity in 123 aa overlap"
        /codon_start=1

```

misc\_feature  
92. .460  
/gene="rplN"  
/note="Pfam match to entry PF00238 Ribosomal L14, Ribosomal protein L14p/L23e, score 277.30, E-value 2e-79"  
misc\_feature  
269. .349  
/gene="rplN"  
/note="PS00049 Ribosomal protein L14 signature"  
474. .788  
/gene="rplX"  
/note="synonym: STY4369"  
474. .788  
/gene="rplX"  
/note="Orthologue of E. coli rplX (RL24 ECOLI); Fastra hit to RL24 ECOLI (103 aa), 99% identity in 103 aa overlap"  
/codon\_start=1  
/transl\_table=1  
/product="50S ribosomal subunit protein L24"  
/protein\_id="CAD09157.1"  
/db\_xref="GI:16505161"  
/db\_xref="GOA:Q821X8"  
/db\_xref="UniProt/TREMBL:Q821X8"  
/translation="MAKIRRDDEVIVLTGDKGRGKGVNVLSSGKVIIVGINLVKK  
HOKVPALNPGGIVKEKALIVSNVAIFNTATGADRVGRFEDGKVRFFKSNSET  
IK"  
486. .656  
/gene="rplX"  
/note="Pfam match to entry PF00467 Ribosomal\_L24, KOW motif, score 89.90, E-value 5.1e-23"  
495. .548  
/gene="rplX"  
/note="PS01108 Ribosomal protein L24 signature"  
803. .1342  
/gene="rplE"  
/note="synonym: STY4370"  
803. .1342  
/gene="rplE"  
/note="Orthologue of E. coli rplE (RL5 ECOLI); Fastra hit to RL5 ECOLI (178 aa), 99% identity in 178 aa overlap"  
/codon\_start=1  
/transl\_table=1  
/product="50S ribosomal subunit protein L5"  
/protein\_id="CAD09158.1"  
/db\_xref="GI:16505162"  
/db\_xref="GOA:P37436"  
/db\_xref="UniProt/Swiss-Prot:P37436"  
/translation="MAKLHDYKDEVNKNLTFENYNSVMQVPRVEKTLTMNGVEAL  
ADKKLDNAADLTAISGQPLITARSSVAFKIRGQPIGCKVTLGERMWEFFER  
LITIAVPRIIRDFRGLSAXSPDGRGNVSGVBRQIIFPEIDYDKDVRVGLDITITTTA  
KSDREGRALIAFDFFPRK"  
872. .1042  
/gene="rplE"  
/note="Pfam match to entry PF00281 Ribosomal\_L5, Ribosomal protein L5, score 120.00, E-value 4.5e-32"  
971. .1021  
/gene="rplE"  
/note="PS00358 Ribosomal protein L5 signature"  
1052. .1336  
/gene="rplE"  
/note="Pfam match to entry PF00673 Ribosomal\_L5\_C, ribosomal L5p family C-terminus, score 205.50, E-value 8.2e-58"  
1280. .1333  
/gene="rplE"  
/note="PS01156 TonB-dependent receptor proteins signature

gene  
2<sup>2</sup>  
1357. .1662  
/gene="rpsH"  
/note="synonym: STY4371"  
1357. .1662  
/gene="rpsH"  
/note="Orthologue of E. coli rpsH (RS14 ECOLI); Fastra hit to RS14 ECOLI (100 aa), 98% identity in 100 aa overlap"  
/codon\_start=1  
/transl\_table=1  
/product="30S ribosomal subunit protein S14"  
/protein\_id="CAD09159.1"  
/db\_xref="GI:16505163"  
/db\_xref="GOA:Q8XEM0"  
/db\_xref="UniProt/Swiss-Prot:Q8XEM0"  
/translation="MAQSKMAREKRVKVALADKYFARAEIKAIISDVNAITDEDRNA  
VLKIQTLPRDSSPSRQRRRCQGTGRPHAFIRKFGLSRIKYEAAMRGEIDGKKASW"  
1357. .1659  
/gene="rpsH"  
/note="Pfam match to entry PF00253 Ribosomal\_S14, Ribosomal protein S14p/S29e, score 177.60, E-value 1.3e-53"  
1543. .1611  
/gene="rpsH"  
/note="PS00527 Ribosomal protein S14 signature"  
1696. .2088  
/gene="rpsH"  
/note="synonym: STY4372"  
1696. .2088  
/gene="rpsH"  
/note="Orthologue of E. coli rpsH (RS8 ECOLI); Fastra hit to RS8 ECOLI (129 aa), 100% identity in 129 aa overlap"  
/codon\_start=1  
/transl\_table=1  
/product="30S ribosomal subunit protein S8"  
/protein\_id="CAD09160.1"  
/db\_xref="GI:16505164"  
/db\_xref="GOA:P02361"  
/db\_xref="UniProt/Swiss-Prot:P02361"  
/translation="TMSQDPIADMLTRIRNGQANKAATMPSSKLRVAIANVKEEG  
FIEDPKVGDTEKLELTLYKFGKAVESIQVSRPGLRIYKKDELPKVMAGIGIA  
VSTRSGVMTDRARQAGLGEIICVYA"  
1708. .2085  
/gene="rpsH"  
/note="Pfam match to entry PF00410 Ribosomal\_S8, Ribosomal protein S8, score 248.20, E-value 3.6e-75"  
1993. .2046  
/gene="rpsH"  
/note="PS00053 Ribosomal protein S8 signature"  
2101. .2634  
/gene="rplF"  
/note="synonym: STY4373"  
2101. .2634  
/gene="rplF"  
/note="Orthologue of E. coli rplF (RL6 ECOLI); Fastra hit to RL6 ECOLI (176 aa), 99% identity in 176 aa overlap"  
/codon\_start=1  
/transl\_table=1  
/product="50S ribosomal subunit protein L6"  
/protein\_id="CAD09161.1"  
/db\_xref="GI:16505165"  
/db\_xref="GOA:Q8XFL6"  
/db\_xref="UniProt/Swiss-Prot:Q8XFL6"  
/translation="MSRVAKAPVVPAVDVINKINGQVITIKNGSELTTLTDAVEVK  
HADNALFGPRDGYADQMAQGTARALNSVIVGTGFTKKQLGVGVRAAVKGVN  
VNLISGSHPVDDHLPAGITAECPYREIYVKGADKQVIGVADLAAVRRPEEYKKG  
GVRYADEVRYTKAKKK"  
2131. .2631  
/gene="rplF"  
/note="Pfam match to entry PF00347 Ribosomal\_L6, Ribosomal protein L6, score 299.10, E-value 5.4e-86"  
2560. .2586  
/gene="rplF"

[illegible]

Salmonella typhi CT18"  
/codon start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA071518.1"  
/db\_xref="GI:29139954"  
/translation="MODITMEARLALESRLAFQBITIEILNTVTVAHEMEXALRNH  
IRLITELKASQPSINASQABETPPPH"  
1293..2111  
/gene="fkpa"  
/locus\_tag="t4052"  
1293..2111  
/gene="fkpa"  
/locus\_tag="t4052"  
/note="corresponds to STY4345 from Accession AL513382:  
Salmonella typhi CT18"  
/codon start=1  
/transl\_table=11  
/product="FMRP-type peptidyl-prolyl isomerase"  
/protein\_id="AA071518.1"  
/db\_xref="GI:29139955"  
/translation="MKSLFKATLLATTMAVAMHVITPAADAAKPAATADSKAFKND  
DOKAAVLAGASIGRYMENSLEKEOKLGIKLDQOLIAGVODAFADKSKLSPQELQRT  
OFFEARYAGASIGRYMENSLEKEOKLGIKLDQOLIAGVODAFADKSKLSPQELQRT  
PMDSTVYVNVKGLTIDKEFPDNSVTGEPSLPLDVIPEWTEGLKNIKKSGKIKLV  
IPPALAYKGVCPGIPANSTIVFDVELLDIKPAKADAKPADADAKAADAANK"  
2388..3110  
/gene="yheo"  
/locus\_tag="t4053"  
2388..3110  
/gene="yheo"  
/locus\_tag="t4053"  
/note="corresponds to STY4346 from Accession AL513382:  
Salmonella typhi CT18"  
/codon start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA071520.1"  
/db\_xref="GI:29139956"  
/translation="MSRSLTNETSELDLDQRPPEQTDPLDKSYEAVVQGLAMLIG  
SHCEIVASLDLQDKSAIRLANGHGTGKIGSPITDIALRLHMTGADSSVSCPTT  
RAKSGVLMKSLITLIRNEQRVIGLTCINMMLDVPSQIMNTFIPPTPEGSAVNPA  
RSVEVLVQTEFLTEFVNADRNVSNNAKRQIVLNTKGIPIKQAINQVADRNLNI  
SKHYLVYIRQKSGDFQGDK"  
3110..3496  
/gene="yhen"  
/locus\_tag="t4054"  
3110..3496  
/gene="yhen"  
/locus\_tag="t4054"  
/locus\_tag="t4054"  
/note="corresponds to STY4347 from Accession AL513382:  
Salmonella typhi CT18"  
/codon start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA071521.1"  
/db\_xref="GI:29139957"  
/translation="MRVAIVTGGAYGTOQASALQAPAHALNHEGHASVFPFREGV  
YNAHLTSPADSDVLAIVWOKLNTQHGVALNICVAALALRGIIIDTEAGRIELPSAN  
LQPGTSLGCLALASLTCRVQF"  
3496..3852  
/gene="yhem"  
/locus\_tag="t4055"  
3496..3852  
/gene="yhem"  
/locus\_tag="t4055"  
/locus\_tag="t4055"  
/note="corresponds to STY4348 from Accession AL513382:  
Salmonella typhi CT18"  
/codon start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA071522.1"  
/db\_xref="GI:29139958"

translation="MKRIAFVSTPAPHSAGREGLDALLATSLATXGVFPISDGV  
FOLLPGQPDVAVLADYIATFKLFDLYIDOCWICAMSLRERGLNANFVNATPLEP  
VALRRELGNVILRF"  
/gene="yhel"  
3860.4147  
/locus\_tag="t4056"  
3860.4147  
/gene="yhel"  
/locus\_tag="t4056"  
/note="corresponds to STY4350 from Accession AL513382:  
Salmonella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAO71523.1"  
/db\_xref="GI:29139959"  
/translation="MLHTLPYHCASGVDPALLRLKEGDALLLLDQVTVAIEGNRFL  
ESLRAPTYVALKEDIDARGIGQISDSVVRVDTFRVRLTVKANQMAN"  
4273.4647  
/gene="tpsl"  
/locus\_tag="t4057"  
4273.4647  
/gene="tpsl"  
/locus\_tag="t4057"  
/note="corresponds to STY4350 from Accession AL513382:  
Salmonella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="30S ribosomal subunit protein S12"  
/protein\_id="AAO71524.1"  
/db\_xref="GI:29139960"  
/translation="MATVQIVRKPRARKAVKSNVPALEACPQKRGVCTRYTTTPRK  
PNSALRKQCVRLTNGFEVTSYIGEGHNLDEHSVILRGGRVXDLPGVRVHTVAGAL  
DCSGVKKROAKSKYGRPKA"  
4743.5213  
/gene="tpsg"  
/locus\_tag="t4058"  
4743.5213  
/gene="tpsg"  
/locus\_tag="t4058"  
/note="corresponds to STY4351 from Accession AL513382:  
Salmonella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="30S ribosomal subunit protein S7"  
/protein\_id="AAO71525.1"  
/db\_xref="GI:29139961"  
/translation="MRRRVITGQKILDPKFGSILAKFVNILMVDGKKSIVASIV  
SALETLAQSGKSELEAFVLENVRPVVEKSRVVGSTVQVPEVPRNLAAMR  
WIVEARRKGDSMALRLANELSDAADNKGTVAKREDVHMAEANKAFAHYRW"  
5310.7424  
/gene="fusa"  
/locus\_tag="t4059"  
5310.7424  
/gene="fusa"  
/locus\_tag="t4059"  
/note="corresponds to STY4352 from Accession AL513382:  
Salmonella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="elongation factor G"  
/protein\_id="AAO71526.1"  
/db\_xref="GI:29139962"  
/translation="MARTPIARVYNIGTSAHIDAGKTTTTERILFYGVNKHIGSEVH  
DGAATMDMEQERGITTTSAATTAFNSGAKQYRPHRINIIDTPGHADPTIEVER  
KVLDAVNVYCAVGQVQPSSTVWRQAKRYVPIAFVKNKDRMGANLAKVQGIKT  
RIGANVPVLQALIGABEGFTGVVDLVKMAIMWNAODQVTEYEDIDPADMDLANEW  
HONLIEASAESELMERKLGSELTEEBEIKOALRQVLANEIIIVTCGSAFKNGVQ  
AMLDAVIDLPSGVPAIVKTIARERGRIVQGHANKREIKERAGDIAAALGLDVT  
PRRYSGVNSGDTVLNYSKTAERERGRIVQGHANKREIKERAGDIAAALGLDVT

Query Match 83.3%; Score 20; DB 1; Length 300620;  
Best Local Similarity 77.3%; Pred. No. 1.1e+03;

Matches 17; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 TCCTGTGGCCSGTFSACSCGA 23  
Db 247494 TCCTGTGGCCGCGTACCCGCA 247473  
RESULT 7  
STWAFSAA/c 1203 bp DNA linear BCT 26-APR-1993  
LOCUS S.-griseus afsa gene encoding a possible A-factor biosynthesis  
DEFINITION protein.  
VERSION M24250  
KEYWORDS A-24250.1 GI:153148  
SOURCE Streptomyces griseus  
ORGANISM Streptomyces griseus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycetaceae; Streptomyces.  
REFERENCE 1 (bases 1 to 1203)  
Horiouchi, S., Suzuki, H., Nishiyama, M. and Beppu, T.  
Nucleotide sequence and transcriptional analysis of the  
Streptomyces griseus gene (afsa) responsible for A-factor  
biosynthesis  
JOURNAL J. Bacteriol. 171 (2), 1206-1210 (1989)  
MEDLINE 89123125  
COMMENT 2492509  
FEATURES  
source location/Qualifiers  
1..1203  
/organism="Streptomyces griseus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1911"  
204..1109  
/note="afsa protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA26693.1"  
/db\_xref="GI:153149"  
/translation="MDAEEVHPVGIEMVHRTPEDEAPRNWVRLGRDPAEVAVL  
HDHPFAPVGGDLDDPLLVAEAMQAMLAFAAGYGLGVHFLTEIDYVCHENLG  
VGRPRTIGVSGFCSDLKMRAGLPAGRGVGVAVRGRGLATGVAATRFSTPKAYRM  
RGDVPVIGISLPTAPVPDAPAGARAEVDVLTSGTGRGVAELVDVDRHPTLPORPD  
HVPEMLLEAKQAACLVAGPAGIVPEARTFRIRYSFSGPCWIGAVVQGADEDTY  
TVRTGHDGDETVSTVLSGPRAG"

Query Match 81.7%; Score 19.6; DB 1; Length 1203;  
Best Local Similarity 79.2%; Pred. No. 3.5e+03;  
Matches 19; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GTCCGTGTGGCCSGTFSACSCGAC 24  
Db 1058 GTCCGTGTGGCCGTGTACCCGAC 1035  
RESULT 8  
AB011413/c 12070 bp DNA linear BCT 07-AUG-1998  
LOCUS Streptomyces griseus genes for Orf2, Orf3, Orf4, Orf5, AfsA, Orf8,  
DEFINITION partial and complete cds.  
AB011413  
ACCESSION AB011413.1 GI:3401946  
VERSION Orf8; AfsA; Orf5; Orf4; Orf3; Orf2.  
KEYWORDS Streptomyces griseus  
SOURCE Streptomyces griseus  
ORGANISM Streptomyces griseus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycetaceae; Streptomyces.  
REFERENCE 1 (sites)  
Uneyama, T.  
AUTHORS Open reading frame encoded around afsA gene  
TITLE Unpublished  
JOURNAL 2 (bases 1 to 12070)  
REFERENCE



FEATURES	
source	Location/Qualifiers
	1..3118
	/organism="Oryza sativa (japonica cultivar-group)"
	/mol_type="mRNA"
	/cultiivar="Nipponbare"
	/db_xref="taxon:39947"
	/clone="002-164-F01"
ORIGIN	
Query Match	80.0%; Score 19.2; DB 8;
Best local Similarity	75.0%; Pred. No. 4.3e+03;
Matches 18; Conservative	4; Indels 0; Gaps 0;
Oy	1 GTCCCTGCTGGCCSGTSAACGCGSAC 24
	:     :     :
Db	1442 GTCCCTGACAGCGCGGTGACACGAC 1465
RESULT 10	
SC0007731	
LOCUS	SC0007731 6727 bp DNA linear BCT 24-AUG-1998
DEFINITION	Streptomyces coelicolor scdR gene, scdA gene, ORFs A,B,X & Z.
ACCESSION	AJ007731
VERSION	AJ007731.1 GI:3425857
KEYWORDS	gamma-butyrolactone binding protein; scdA gene; scdR gene.

[illegible]

```

CDS
3032..3679
/gene="gcbR"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="gamma-butyrolactone binding protein"
/protein_id="CAA07628.1"
/db_xref="GI:3425861"
/db_xref="GCA:086852"
/db_xref="UniProt/TREMBL:O86852"
/translation="MAKODRAIRTRQITLIDAAAVFEKQYQAAITTEILKVAQVTKG
ALYHFQSGKEIALGVFPAQBPQVAFQPLQELIDMGVLFGRLEINVAAGVRA
LSMQOAGHLDGRGPRFRMHTLKLNOAKENBELPHVVTTSADIYVTFPGIQ"
VSQYVSDYQDLHENRYALKQITLPAIVPSTLALDLSEEGALAAELATGCD"
complement(3795..4529)
/gene="OrfX"
/complement(3795..4529)
/gene="OrfX"
/codon_start=1
/transl_table=11
/product="3-ketoacyl-AcP/CoA reductase"
/protein_id="CAA07629.1"
/db_xref="GI:3425862"
/db_xref="GCA:086853"
/db_xref="UniProt/TREMBL:O86853"
/translation="WTGSSRGTIGRTALRDLADLAIVHYGSSEAAARETVETIRSS
GGQALAIKAEISGVGDAALYAAPDAGKEGVPEPDIIVNNAVSGSGKITVTESE
VFPLAVNVNAPLFLVOHGLKRLDGGRIINISAAIRAPPSIGVAMTKGAVDTL
TLTARQIGERGITVNAVAPGFVETDNNARRRQTPFAAALAANSVFNRIGRPDDIAD
VVAFLASDSSWITGQYVDATGGTLL"
complement(4746..6646)
/gene="OrfZ"
/complement(4746..6646)
/gene="OrfZ"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAO07630.1"
/db_xref="GI:3451475"
/db_xref="UniProt/TREMBL:O86854"
/translation="MMRVRCRGVLGRGRRRRRPAKSRPGAAAPARWSSRPAORPDA
GGPTDARARAVRDPFAAGAPRLPEPDAAEPLDAGQAGTSSAGGATTAHAPAPAS
PGTHABEASPGADTQGSVAAYVGRITVSSAPAKARAPRMAPDRDQDRIFFPAL
PFRALAEGPATWAGNAWVTALEEGALDAALRGRGYAEHGVDAITVPGVLVAV
VGRSGRYRQVVALRTIGDSMDWRFLDAAVERGHLAALDGEPLSLADLADRGVPL
LLPFGDLPAPRCSCEPSGHPCKHAALCYQARLLDADPFVLLLRGSEKALDALS
RNRNABARAADNRGPGPLPGVRAAALTPRALPLPAPLPAPHPROPAPPAAGG
PYPRPLDIALDAARPHALLTGRGADVPVGLTLMODAVRLAARPGSGTATGATYLA
SLAAGAGDTLELPAVAANRQGLAGLDVIEEWMDFPAGFDRAPMTLADADIPARF
PMNRRLTPHGRHVQLRIGRTGLWYAYESELGRBDMWPGTDDLPVATLGLGGPGDP"

ORIGIN
Query Match      80.0%; Score 19.2; DB 1; Length 6727;
Blast Local Similarity 75.0%; Pred. No. 3.9e+03;
Matches 18; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Oy      1  GTCCGTGTCGCCGTGATCGSAC 24
|||||  |||||  |||||  |||||  |||||
Db      2021  GTCTGTGATGCCCGGTGACCCGCAC 2044

RESULT 11
LOCUS      AP003954      131711 bp      DNA      linear      HTG 21-MAR-2002
DEFINITION      Oryza sativa (japonica cultivar-group) chromosome 7 clone
SOURCE      OJ181.B07, *** SEQUENCING IN PROGRESS ***.
VERSION      AP003954
KEYWORDS      HTG; HTGS PHASE2.
ORIGIN      Oryza sativa (japonica cultivar-group)
            Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source  
ORIGIN  
Query Match  
Best Local Similarity  
Matches  
Db  
oy  
RESULT 12  
AP004269  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1  
Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.

1  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Oryza sativa nipponbare (Ga3) genomic DNA, chromosome 7, BAC  
clone: OJ1381\_B07  
Published Only in Database (2001)  
2 (bases 1 to 131711)  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Direct Submission  
Submitted (25-JUL-2001) Takuji Sasaki, National Institute of  
Agrobiological Resources, Rice Genome Research Program; Kamondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail: tsasa@nias.affrc.go.jp, URL: http://rgrp.dna.affrc.go.jp/,  
Tel: 81-298-38-7441, Fax: 81-298-38-7468)  
The nucleotide sequence of this BAC clone was generated by  
combining Monsanto and RGP-Japan sequencing data.  
NOTE: It currently consists of 1 contigs. Gaps between the contigs  
are represented as runs of N. The order of the pieces is believed  
to be correct as given, however the sizes of the gaps between them  
are based on estimates that have provided by the submitter. This  
sequence will be replaced by the finished sequence as soon as it is  
available and the accession number will be preserved.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
Location/Qualifiers  
1..131711  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/chromosome="7"  
/clone="OJ1381\_B07"

80.0%; Score 19.2; DB 2; Length 131711;  
75.0%; Pred. No. 2.5e+03;  
4; Mismatches 2; Indels 0; Gaps 0;

95743 GTCCTGACGCGCGGTGACACGCAC 95720

1 GTCCGTGGCGCGGTGACACGCAC 24  
|||||:|||||:|||||:|||||  
GTCCCTGACGCGCGGTGACACGCAC 95720

145772 bp DNA linear PLN 22-JUL-2004  
Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 7,  
PAC clone: P0048D08.  
AP004269  
AP004269.4 GI:50509267

Oryza sativa (japonica cultivar-group)  
Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.

1  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Oryza sativa nipponbare (Ga3) genomic DNA, chromosome 7, PAC  
clone: P0048D08  
Published Only in Database (2001)  
2 (bases 1 to 145772)  
Sasaki, T., Matsumoto, T. and Yamamoto, K.  
Direct Submission  
Submitted (17-OCT-2001) Takuji Sasaki, National Institute of  
Agrobiological Sciences, Rice Genome Research Program; Kamondai  
2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
(E-mail: tsasa@nias.affrc.go.jp, URL: http://rgrp.dna.affrc.go.jp/,  
Tel: 81-298-38-7441, Fax: 81-298-38-7468)  
On Jul 22, 2004 this sequence version replaced gi:23616992.



Genes were predicted from the integrated results of the following:

GENSCAN (<http://CCR-081.mt.edu/genSCAN.html>), FGENESH (<http://www.softberry.com/>), Genemark.hmm (<http://opal.biology.gatech.edu/Genemark/>), GlimmerM ([http://www.tigr.org/tcd/glimmer/glmr\\_form.html](http://www.tigr.org/tcd/glimmer/glmr_form.html)), RiceHMM (<http://xrp.dna.affrc.go.jp/RiceHMM/>), SplicePredictor (<http://bioinformatics.iastate.edu/cgi-bin/sp.cgi>), sim4 (<http://glabin.cse.psu.edu/html/docs/sim4.html>), gap2 (<http://www.tigr.org/software/glimmer/>), BLASTn and BLASTx. The genomic sequence was searched against NCBI Nonredundant Protein database, nr (<ftp://ncbi.nlm.nih.gov/blast/db>) and the cDNA sequence database at RGP or DDBJ. Protein homologies of the coding regions were searched against NCBI Nonredundant Protein database with BLASTp. ESTs represent the identified cDNA sequences using BLASTn with the corresponding DDBJ accession no. and RGP clone ID. Full-length cDNAs represent the identified cDNA sequences using BLASTn with the corresponding DDBJ accession no.

A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with full-length cDNA or EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.

The orientation of the sequence is from T7 to SP6 of the PAC clone. This sequence of P0048D08 clone has an overlap with OSJNB0041106 (DDBJ: AP005176) clone at 5' end and with P0650C03 (DDBJ: AP005320) at 3' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://xrp.dna.affrc.go.jp/Genomeseg.html>.

FEATURES

SOURCE

1.145772  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/chromosome="7"  
/clone="P0048D08"  
5733..8129  
/gene="P0048D08.101"  
join(<5733..5870,6406..6555,6829..6905,7181..7504,  
7724..7765,7865..7964,8064..>8129)  
/gene="P0048D08.101"  
/note="start and end point are not identified"  
join(5733..5870,6406..6555,6829..6905,7181..7504,  
7724..7765,7865..7964,8064..8129)  
/gene="P0048D08.101"  
/note="similar to Oryza sativa chromosome 1,  
OSJNB0091E23.21"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="BAD30555.1"  
/db\_xref="GI:50509268"  
/translation="MEDGKSVLPARLRPSNVTVPITNORSTOAAHDSSENKRCV  
TDELSITVYGRVHSTIENLDKGVAPNLPPTDLSHSTGHRWTRKDAISH  
TASIGAREIILANRPSNRGRVTTAMSTNDVASLPLPLQVVELALLSPRATY  
RFSHFHFKMSYDCLVDDHSHVSPTTSISGVTTALSTSDVASLPLPLQV  
ELRLPLGPRPAIDFSLYLFNFKVDPRPNRLDLGTPRTDKALSDVEKKDHVRS  
RAGQSKSLILAEFRSDKAG3"  
10272..10955  
/gene="P0048D08.102"  
<10272..>10955  
/gene="P0048D08.102"  
/note="start and end point are not identified"  
10272..10955  
/gene="P0048D08.102"  
/note="similar to Oryza sativa chromosome 7, P0048D08.3"  
/codon\_start=1  
/product="hypothetical protein"

/protein\_id="BAD30555.1"  
/db\_xref="GI:50509268"  
/translation="MPRRDCPSRIIDQLITAKLIDINSTDPNSDSAGYKSMGR  
LRHQIQLPILPALSSMALPRSSHISGPIHCHSELAVEAKGVKSPGSG  
SGIPKSSFFKDEVRVTSGLRTGLTALVMWLGAVVSDSDADGHLVYGNP  
NPPDDPQGVRAVVKDVKPARBPATPANMAAPRNKAGKSLPLKMELEKELLR  
ANPALLAS"  
14881..15381  
/gene="P0048D08.103"  
<14881..>15381  
/gene="P0048D08.103"  
/note="start and end point are not identified"  
14881..15381  
/gene="P0048D08.103"  
/note="similar to Oryza sativa chromosome 7, P0048D08.3"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="BAD30557.1"  
/db\_xref="GI:50509270"  
/translation="MALPRSSPRSSRPITIRAVDAKGRVSKPRKSGPYKSSAP  
STAVPSFKGDEVRVTPVGRIGTTLRLVMWLGAVVSDSDADGHLVYGNP  
RDDEFRVRAVVKDVKPARBPATPANMAAPRNKAGKSLPLKMELEKELLR  
EALPAS"  
complement(16394..21283)  
/gene="P0048D08.104"  
complement(join(<16394..16504,17089..17222,17771..17894,  
17988..18257,18967..19161,19242..19334,19417..19545,  
19652..>21283))  
/gene="P0048D08.104"  
/note="start and end point are not identified"  
complement(join(<16394..16504,17089..17222,17771..17894,  
17988..18257,18967..19161,19242..19334,19417..19545,  
19652..>21283))  
/gene="P0048D08.104"  
/note="predicted by FGENESH etc."  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="BAD30558.1"  
/db\_xref="GI:50509271"  
/translation="MAGVVGSKSNRDMSDCDNAFREAGRAIVALTDEEGCHAGL  
SLATLVGRYPWSPILARAILARCYLQNSRQOEVOLELAIVLAPRCPHIALIDIL  
ISMDLPDEABVRBALRVAPEDPALHYFVSRSYSSADHNDPFLERKAGRET  
IRQGRARIRBKGGQAASPRPASPREPPTVDVIGADWSRMSSEERALKVSG  
EMKSYCRSGMDMTSMUSDNEVFKAGSCPPFSGNITYEFAFKSHIDBEHVGKE  
FLSLVPSRISERSELLRSRWEPETDDDLAGRTKILREKELVFEILDLEVSINL  
YIMKFTMNRVPAPLVVMSCSGCGIGSLTLOLCELYPLKLVVOTORGEHQ  
KHNHDSOODSLVHTHRCNHNKRNESQESQDSLVGITSQETGSLSPCEKI  
ASRPTDSSQADRLFACILSEPLEEDMELCFMSWRCFPDGGPILNINIRALGAKL  
KFSWELEKIGQGVFLPKAIPERDDITITYPSWIGSARVELLIDAYDVYKEEL  
LKTCQVCLAVISPIAKACLMAKLVNDPLEDALAHQONCHRPQVPIADILRSIMHR  
RFGCDLWEIPCIISPDVARVRAVILIRIFRSMDCQCDLPSXAFVNDLSRSLVDE  
KAGNISAVRVESILERLHAQTLIFEPKESLVPOTAVPSLACICLAHNFGLH  
IIEKKCNVNEVPMKSTFPFHSINLSGVSGETLSESELLKAVDKOSVCDFRNGCG  
HRIIRLYMYPHPFMYIVLRMPDNKGNINNHKULISLAELDISHIEGQSESMYTL  
VSAQESWKGILIDRYSQANLRPELIFEHGRKRDHRLLL"  
21456..21816  
/gene="P0048D08.105"  
join(<21456..21518,21709..>21816)  
/gene="P0048D08.105"  
/note="start and end point are not identified"  
join(21456..21518,21709..21816)  
/gene="P0048D08.105"  
/note="predicted by Genemark.hmm etc."  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="BAD30559.1"  
/db\_xref="GI:50509272"  
/translation="MPRRRTKVAEADEBAGBQKLLFTALTACADERKKGSKYMBE  
KGGKGGKGGEL"  
complement(22480..24657)  
/gene="P0048D08.106"  
complement(join(<22480..22595,22769..22915,23074..23122,  
23246..23371,23479..>24657))



On or before Oct 30, 2002 this sequence version replaced  
gi:20520661, gi:20520749, gi:20520754, gi:20520755, gi:20520688,  
gi:20520689, gi:20520815, gi:20520603, gi:20520917, gi:20520666,  
gi:20520864.

```

FEATURES
    source          location/Qualifiers
1. .290850
   /organism="Streptomyces coelicolor A3(2)"
   /mol_type="genomic DNA"
   /strain="A3(2)"
   /db_xref="taxon:100226"
97. .1101
   /gene="SC06257"
   /note="synonym: SCAH10.22"
97. .1101
   /gene="SC06257"
   /note="SCAH10.22, probable ABC transport system sugar
binding lipoprotein, len: 334 aa; similar to TR:CA841563
(EMBL:U049727) Streptomyces coelicolor putative secreted
solute binding protein, 337 aa; fasta scores: opt: 1003
z-score: 1130.9 E(): 0; 49.8% identity in 325 aa overlap,
to SW:RBSB_BACSU (EMBL:Z92953) Bacillus subtilis
D-ribose-binding protein precursor RBSB, 305 aa; fasta
scores: opt: 206 z-score: 238.7 E(): 6.3e-06; 23.0%
identity in 235 aa overlap and to TR:CA841563 (EMBL:SC9B1)
Streptomyces coelicolor SC9B1, 17c, 337 aa; fasta scores:
opt: 1003 z-score: 1045.2 E(): 0; 49.8% identity in 325 aa
overlap. Contains match to Pfam entry PF00532
Peripla_BP_like, Periplasmic binding proteins and Lacti
family and a match situated in the correct position to
Prosite entry PS00013 Prokaryotic membrane lipoprotein
lipid attachment site. Contains also possible N-terminal
region signal peptide sequence"
   /codon_start=1
   /transl_table=11
   /product="putative ABC transport system sugar binding
lipoprotein"
   /protein_id="CA860176.1"
   /db_xref="GI:6273662"
   /db_xref="UniProt/TREMBL:O9RKT5"
   /translation="MARFTVGVIALAGALSVSLAGCSSYTGKRAEDARKAASAEGRRA
AVDPRTFAMITHSGDGFPMIDVOSGAOAIVKQINIFLSHDAAGOOAALVNAASA
IDKKVGIITVLAKPDMKSAALRAHKAQIPVITVNSGESSEKRGALFVGGDEETRA
GEAVGERLNERGRQAVCVLHEGQNVNBEQSCGVEPTDGRKQRLVYNTSNPDVQS
AIEAKGQTDKSDVAVVTLGAPVADTVAKAQGASKKEIDTFDINAAVAGLADGTGLG
FADVQGPYLQGYEAVDILMLVKNYNAVLTGGGRPVLTPGPITITDDAALADLYTKRGTR
"
133. .165
   /gene="SC06257"
   /note="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
241. .966
   /gene="SC06257"
   /note="Pfam match to entry PF00532 Peripla_BP_like,
Periplasmic binding proteins and Lacti family., score
13.80, E-value 2.1e-06"
1098. .2138
   /gene="SC06258"
   /note="synonym: SCAH10.23"
1098. .2138
   /gene="SC06258"
   /note="SCAH10.23, probable ABC transport system sugar
permease, len: 346 aa; similar to TR:O68120
(EMBL:AF010496) Rhodococcus capsulatus ribose transport
system permease protein RbsC; fasta scores: opt: 445
z-score: 508.0 E(): 6.4e-21; 30.5% identity in 351 aa
overlap and to SW:RASC_ECOLI (EMBL:U10328) Escherichia
coli ribose transport system permease protein RbsC, len:
321 aa; fasta scores: opt: 286 z-score: 329.9 E():
5.3e-11; 30.1% identity in 329 aa overlap. Contains
possible hydrophobic membrane spanning regions"
   /codon_start=1
   /transl_table=11
   /product="putative ABC transport system sugar permease"

```

/protein\_id="CAB60177.1"  
/db\_xref="GI:6273663"  
/db\_xref="GI:6273665"  
/db\_xref="GOA:Q9RKT4"  
/translation="MNATOVKADDERIIQTSRLRLGLGRPEIGSVGAIAVFEFAF  
ADSLFHAASLSTLVLAFTIGMAVPAVLLMIGFEDISAGMTWSLVSMSYOM  
TANWVGVSILYLTIALGAPNGFWLPTKLSPIITIGTFLM.TGNLITGLVDGT  
VSTKSIADMEGPPSAQDVPAFTITTTGGVFKYTIIMWLAIVAVSMILRTKAGWIF  
AVGNKDAARAVGVPAKTKI GLYVGFGAMI SGQHLFSYDVVQSGEGVNELIYL  
IAAVGGCLITGCGYSAVSAVGAFIPGWTSGIYFAEWBDFWPFPLGAMILLATLL  
NANWKRAREATK"  
2124. 2127  
/gene="SC06258"  
2135. 2926  
/gene="SC06259"  
/note="synonym: SCAH10.24"  
2135. 2926  
/gene="SC06259"  
/note="SCAH10.24, probable ABC sugar transport ATP binding  
protein, len: 263 aa; similar to TR:068121 (EMBL:AF010496)  
Rhodobacter capsulatus ribose transport ATP-binding  
protein, 305 aa; fasta scores: opt: 612 z-score: 706.1  
E(): 5.9e-32; 45.6% identity in 217 aa overlap, to  
SW:RBSA.ECOLI (EMBL:M1169) Escherichia coli; ribose  
transport ATP-binding protein RbsA, 501 aa; fasta scores:  
opt: 509 z-score: 585.6 E(): 3e-25; 31.5% identity in 254  
aa overlap and to TR:CA15787 (EMBL:SC7B7) Streptomyces  
coelicolor str787.07, 260 aa; fasta scores: opt: 592  
z-score: 521.2 E(): 1e-23; 43.1% identity in 246 aa  
overlap. Contains Pfam match to entry PF00005 ABC tran.  
ABC transporter and Prosite PS00017 ATP/GTP-binding site  
motif A (P-loop) and PS00211 ABC transporters family  
signature"  
/codon\_start=1  
/transl\_table=11  
/product="probable ABC sugar transport ATP binding  
protein"  
/protein\_id="CAB60178.1"  
/db\_xref="GI:6273664"  
/db\_xref="GOA:Q9RKT3"  
/db\_xref="UniProt/TREMBL:Q9RKT3"  
/translation="MTBMBDDRRLVLEISGVSNYGNVAVLALGVSLVFNAGRTCVLGD  
NAGKSTIKITAGLHNDGCTSLDGETRISLSPRRALDRIATVYDILAVPLMPY  
WNPFLLGSEPRGVAPFRMDVDHNRRTTHALLMGIDLDDVDPIGLTSGGEKQCV  
ALARAIVGAKTLVLDDEPTALGVQSGVAVLYVAARDQGLGVLLITHNPHAYLIG  
DRFVLKRGITWVAQTRDVEVTLDELQMGAGSTEDDDRHLERLG"  
2240. 2815  
/gene="SC06259"  
/note="Pfam match to entry PF00005 ABC tran, ABC  
transporter, score 168.70, E-value 9.6e-47"  
2261. 2284  
/gene="SC06259"  
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
2588. 2632  
/gene="SC06259"  
/note="PS00211 ABC transporters family signature"  
3042. 4190  
/gene="SC06260"  
/note="synonym: SCAH10.25"  
3042. 4190  
/gene="SC06260"  
/note="SCAH10.25, possible sugar kinase, len: 382 aa;  
similar to SW:GLK\_STRCO (EMBL:X65932) Streptomyces  
coelicolor glucokinase (EC 2.7.1.2) (glucose kinase) GLK,  
317 aa; fasta scores: opt: 1022 z-score: 1142.9 E(): 0;  
46.6% identity in 311 aa overlap and to TR:CAB51974  
EMBL:SC6B10 Streptomyces coelicolor SC6B10.20C, 317 aa;  
fasta scores: opt: 1022 z-score: 1035.4 E(): 0; 46.6%  
identity in 311 aa overlap. Contains a match to Pfam entry  
PF00480 ROK. ROK family and Prosite entry PS01125 ROK  
family signature"  
/codon\_start=1  
/transl\_table=11

/product="putative sugar kinase"  
/protein\_id="CAB60179.1"  
/db\_xref="GI:6273665"  
/db\_xref="GOA:Q9RKT2"  
/db\_xref="UniProt/TREMBL:Q9RKT2"  
/translation="MSTYRDPFAIGSRRAVLTGTGRRBRSHLTAPRPVTVGIDIG  
GTFKMAGVADAGNILLKRLRETPDKRSKRVVEDTLVETLDSDDHVAVIGAA  
GMVADADRRLVFAHLSWRNBPRLDRLAGSLAVVYDNDANTAAEMRGARGSD  
HLVWITLTGGA.ILEDGOVKRKIVVAFSGHMQVPPGGRCPGNGCBQISGS  
NALVREARELAAADS PVA YGI IERVKSIDITGPMITTELRBDNACVELLDITIGW  
LGVGIANLAAALDPSCFVIGGVSAADDLIGPARDAFKRDLTGRTGYRPAIRVRAOL  
GPEAGVGAADLSRLVARRPRRAKRRVERRYERYAARRESBSL"  
3159. 3704  
/gene="SC06260"  
/note="Pfam match to entry PF00480 ROK, ROK family, score  
215.50, E-value 5.4e-63"  
3540. 3623  
/gene="SC06260"  
/note="PS01125 ROK family signature"  
4244. 4247  
4259. 4768  
/gene="SC06261"  
Query Match 80.0%; Score 19.2; DB 1; Length 290850;  
Best Local Similarity 75.0%; Pred. No. 2.2e+03;  
Matches 18; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GTCTGTGTGCGCGTSACSCGSAC 24  
Db 12986 GTCTGTGATGCCCGGTGACCCGCAC 12963  
RESULT 14  
BC032284/c 1851 bp mRNA linear ROD 29-JUN-2004  
LOCUS BC032284  
DEFINITION Mus musculus porcupine homolog (Drosophila), transcript variant  
Mporc-c, mRNA (CDNA clone MGC:40733 IMAGE:5362485), complete cds.  
ACCESSION BC032284  
VERSION BC032284.1 GI:21595487  
KEYWORDS MGC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1851)  
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,  
Klausner,R.D., Collins,F.S., Wagner,J., Shenmen,C.M., Schuler,G.D.,  
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,  
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heide,F.,  
Dietchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,  
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,  
Schaefer,T.E., Brownstein,M.J., Umed,T.B., Toshiyuki,S.,  
Carninci,P., Prange,C., Raha,S.S., Logguello,N.A., Peters,G.U.,  
Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,  
McKernan,K.U., Malek,U.A., Gunaratne,P.H., Richards,S.,  
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.U., Hulyk,S.W.,  
Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,  
Fahey,J., Heltón,E., Kettelman,M., Madan,A., Rodriguez,S.,  
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,  
Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D.,  
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,  
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalins,D.E.,  
Scherer,A., Schein,J.E., Jones,S.J., and Marra,M.A.  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2 (bases 1 to 1851)  
Strausberg,R.  
Direct Submission  
Submitted (06-JUN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

REMARK  
COMMENT

USA  
NHL-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-rc@mail.nih.gov](mailto:cgapbs-rc@mail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)  
Aklter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Grant, S., Guan, X., Gupta, J., Haghighi, P.,  
Hansen, N., Ho, S.-L., Karlins, B., Kwong, P., Latic, P., Legaspi, R.,  
Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Stantrop, S., Thomas, P.J., Touchman, J.W.,  
Turgerson, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
Series: IRAK Plate: 65 Row: d Column: 4  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 22094150.

FEATURES

source

1..1851  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="MGC:40733 IMAGE:5362485"  
/cissue\_type="Eye, retina, mouse strain C57Bl/6"  
/clone\_lib="NHL MGC\_94"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1..1851  
/gene="porcn"  
/note="synonyms: Mporc-a, Mporc-b, Mporc-c, Mporc-d, Ppn,  
Mporc, porc"  
/db\_xref="LOCUSID:53627"  
/db\_xref="GI:21595488"  
144..1511  
/gene="Porcn"  
/product="porcupine homolog"  
/protein\_id="AA032284.1"  
/db\_xref="GI:21595488"  
/db\_xref="LOCUSID:53627"  
/db\_xref="MGI:1890212"

CDS

/product="porcupine homolog"  
/protein\_id="AA032284.1"  
/db\_xref="GI:21595488"  
/db\_xref="LOCUSID:53627"  
/db\_xref="MGI:1890212"  
/translation="MATESRQEFPOQLQGLPTVOGQDQIWLITIGPACLLMR  
LGPSYKHAATAVAGFFSLYHFQQLMVAWVSLSLCYLVLFCRSHSRGVLSVT  
ILYILMGEMHMDVTVMHMRGAQOMVAMKAVSLGPDLDGEVAVSPVEFNGYLY  
FVGTIVGPMISFHSYLOAVOGRPLSRMLKVAASLALCLVSTCGVPLPYF  
IPLDGLRLKRGTVRWMLRAVESAVSFHSYFVGLSEATATLGAFTGFEKDLHM  
DLTVSRPLNVELPRSNVETVSMNLPMSYVNTLVVVKVLAALGTFSAVLYVAASALL  
HGSEFHLAIVLISLAFITVYEHVLRKLAQILSACILSRCLPDSCHRRRLGVRALN  
NLFGALAIFFHLSTYGLFVDVDDTTEQSGYGAIVYHKSELSWASHWTFCWATF  
YRLIG"

ORIGIN

Query Match 78.3%; Score 18.8; DB 10; Length 1851;  
Best Local Similarity 70.8%; Pred. No. 6.6e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

LOCUS AB036747 1886 bp mRNA linear ROD 20-JUL-2000

DEFINITION  
AB036747  
AB036747.1 GI:6714507  
porcupine-A.  
porcupine-A.  
Mus musculus (house mouse)  
MUS MUSCULUS  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (sites)  
Tanaka, K., Okabayashi, K., Asashima, M., Perrimon, N. and Kadowaki, T.  
The evolutionarily conserved porcupine gene family is involved in  
the processing of the Wnt family  
Bur. J. Biochem. 267 (13), 4300-4311 (2000)

REFERENCE

JOURNAL  
MEDLINE  
PUBMED  
10866835  
2 (bases 1 to 1886)  
Kadowaki, T. and Tanaka, K.  
Direct Submission  
Submitted (08-JUN-2000) Tatsuhiko Kadowaki, Nagoya University,  
Graduate Program for Regulation of Biological Signals; Chikusa,  
Nagoya, Aichi 464-8601, Japan  
(E-mail: [emi@nagrl.agr.nagoya-u.ac.jp](mailto:emi@nagrl.agr.nagoya-u.ac.jp), Tel:81-52-789-5237)

FEATURES

source

1..1886  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
1..1886  
/gene="Mporc-a"  
174..1526  
/gene="Mporc-a"  
/codon\_start=1  
/product="porcupine-A"  
/protein\_id="BA089468.1"  
/db\_xref="GI:6714508"

gene

CDS

/translation="MATESRQEFPOQLQGLPTVOGQDQIWLITIGPACLLMR  
LGPSYKHAATAVAGFFSLYHFQQLMVAWVSLSLCYLVLFCRSHSRGVLSVT  
ILYILMGEMHMDVTVMHMRGAQOMVAMKAVSLGPDLDGEVAVSPVEFNGYLY  
FVGTIVGPMISFHSYLOAVOGRPLSRMLKVAASLALCLVSTCGVPLPYF  
IPLDGLRLKRGTVRWMLRAVESAVSFHSYFVGLSEATATLGAFTGFEKDLHM  
DLTVSRPLNVELPRSNVETVSMNLPMSYVNTLVVVKVLAALGTFSAVLYVAASALL  
HGSEFHLAIVLISLAFITVYEHVLRKLAQILSACILSRCLPDSCHRRRLGVRALN  
NLFGALAIFFHLSTYGLFVDVDDTTEQSGYGAIVYHKSELSWASHWTFCWATF  
YRLIG"

ORIGIN

Query Match 78.3%; Score 18.8; DB 10; Length 1886;  
Best Local Similarity 70.8%; Pred. No. 6.6e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

LOCUS AB036748 1901 bp mRNA linear ROD 20-JUL-2000

DEFINITION  
AB036748  
AB036748.1 GI:6714509  
porcupine-C.  
porcupine-C.  
Mus musculus (house mouse)  
MUS MUSCULUS  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (sites)  
Tanaka, K., Okabayashi, K., Asashima, M., Perrimon, N. and Kadowaki, T.  
The evolutionarily conserved porcupine gene family is involved in  
the processing of the Wnt family  
Eur. J. Biochem. 267 (13), 4300-4311 (2000)

LOCUS AB036747 1886 bp mRNA linear ROD 20-JUL-2000

REFERENCE 2 (bases 1 to 1901)  
 AUTHORS Kadowaki,T. and Tanaka,K.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-JAN-2000) Tatsuhiko Kadowaki, Nagoya University,  
 Graduate Program for Regulation of Biological Signals, Chikusa,  
 Nagoya, Aichi 464-8601, Japan  
 (E-mail:eml@nagrl.agr.nagoya-u.ac.jp, Tel:81-52-789-5237)

## FEATURES

source

1..1901  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 1..1901  
 /gene="Mporc-c"  
 174..1541  
 /gene="Mporc-c"  
 /codon\_start=1  
 /product="porcupine-C"  
 /protein\_id="BAA89469.1"  
 /db\_xref="GI:6714510"

CDS

gene

## ORIGIN

Query Match 78.3%; Score 18.8; DB 10; Length 1901;  
 Best Local Similarity 70.8%; Pred. No. 6.5e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCTGTGCGCCSGTSACSGSAC 24  
 |||||:||||:||||:||||:||||  
 120 GTCTGTGCGCGCGGAGCGGCAC 97

RESULT 17  
 AB036746/c 1904 bp mRNA linear ROD 20-JUL-2000  
 LOCUS Mus musculus Mporc-b mRNA for porcupine-B, complete cds.  
 AB036746  
 AB036746.1 GI:6714505  
 porcupine-B.  
 Mus musculus (house mouse)  
 SOURCE  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (sites)  
 Tanaka,K., Okabayashi,K., Asashima,M., Perrimon,N. and Kadowaki,T.  
 The evolutionarily conserved porcupine gene family is involved in  
 the processing of the Wnt family  
 Eur. J. Biochem. 267 (13), 4300-4311 (2000)

JOURNAL MEDLINE  
 20325112  
 10866835  
 2 (bases 1 to 1904)  
 Kadowaki,T. and Tanaka,K.  
 Direct Submission  
 Submitted (08-JAN-2000) Tatsuhiko Kadowaki, Nagoya University,  
 Graduate Program for Regulation of Biological Signals, Chikusa,  
 Nagoya, Aichi 464-8601, Japan  
 (E-mail:eml@nagrl.agr.nagoya-u.ac.jp, Tel:81-52-789-5237)

## FEATURES

source

1..1904  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 1..1904  
 /gene="Mporc-b"  
 174..1544

CDS

## ORIGIN

Query Match 78.3%; Score 18.8; DB 10; Length 1904;  
 Best Local Similarity 70.8%; Pred. No. 6.5e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTCTGTGCGCCSGTSACSGSAC 24  
 |||||:||||:||||:||||:||||  
 DB 120 GTCTGTGCGCGCGGAGCGGCAC 97

RESULT 18  
 AB036749/c 1919 bp mRNA linear ROD 20-JUL-2000  
 LOCUS Mus musculus Mporc-d mRNA for porcupine-D, complete cds.  
 AB036749  
 AB036749.1 GI:6714511  
 porcupine-D.  
 Mus musculus (house mouse)  
 SOURCE  
 ORGANISM  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (sites)  
 Tanaka,K., Okabayashi,K., Asashima,M., Perrimon,N. and Kadowaki,T.  
 The evolutionarily conserved porcupine gene family is involved in  
 the processing of the Wnt family  
 Eur. J. Biochem. 267 (13), 4300-4311 (2000)

JOURNAL MEDLINE  
 20325112  
 10866835  
 2 (bases 1 to 1919)  
 Kadowaki,T. and Tanaka,K.  
 Direct Submission  
 Submitted (08-JAN-2000) Tatsuhiko Kadowaki, Nagoya University,  
 Graduate Program for Regulation of Biological Signals, Chikusa,  
 Nagoya, Aichi 464-8601, Japan  
 (E-mail:eml@nagrl.agr.nagoya-u.ac.jp, Tel:81-52-789-5237)

## FEATURES

source

1..1919  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 1..1919  
 /gene="Mporc-d"  
 174..1559  
 /gene="Mporc-d"  
 /codon\_start=1  
 /product="porcupine-D"  
 /protein\_id="BAA89470.1"  
 /db\_xref="GI:6714512"

CDS

gene

1..1919  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 1..1919  
 /gene="Mporc-d"  
 174..1559  
 /gene="Mporc-d"  
 /codon\_start=1  
 /product="porcupine-D"  
 /protein\_id="BAA89470.1"  
 /db\_xref="GI:6714512"  
 /translation="MATESROEFFQOLLQGLLPTVOQGLDQIWLTLTICFACRLMR  
 LGLPSYLKHASTVAGGFPSTLYHFFQLHMVWVVLISLCYVLFCRSHSHRGVFLSVT  
 ILIYLKGMHMDVTYVHKRKGOMIVAKKAVSLGPDLDGEGVAVSPVEPMGYLY  
 FVGITVFGPMISFHSYLOAVQGRPLSRKWLKVARSLALALCLVLTCTGPIPLPYF  
 IPLDGDRLRNKRKARWMLRAYEAVSFHSNFFVGLSEATATLACGFTKEDHLE  
 WDLTVSRPLNVELPRSMVEVVTSMNLPMSYLNWVFKNALRLGTFSAVLVYASAL  
 LHGSFHLAAILAFITVEHVLKRLAQIISACILSKRCIPDCSHRRRLGLGVRL  
 LNLFGALAIPLHLSYGLSLFDVVDVDTTEEGYGMATYVHKMSLSMAHWTFCGWI  
 FYRLIG"

## ORIGIN

Query Match 78.3%; Score 18.8; DB 10; Length 1919;  
 Best Local Similarity 70.8%; Pred. No. 6.5e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCCGTGCGCGCGTSGACGCGSAC 24  
 |||||:||||:||||:||||:||||:  
 Db 120 GTCCGTGCGCGCGCGCGCGCAC 97

RESULT 19  
 SV117268/c 8992 bp DNA linear ECT 02-AUG-2002  
 LOCUS Streptomyces viridochromogenes phbB gene, phc gene, ORF1 and  
 DEFINITION ORF1.  
 Y17268  
 Y17268.2 GI:22095155  
 ORF1; ORF1; phosphinothricin tripeptide synthetase B;  
 phosphinothricin tripeptide synthetase C; phbB gene; phc gene.  
 Streptomyces viridochromogenes  
 Streptomyces viridochromogenes  
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 Streptomycetaceae; Streptomycetaceae; Streptomyces.

SOURCE ORGANISM  
 REFERENCE 1 Schwartz, D., Grammel, N., Reckienwald, J., Keller, U. and Wohleben, W.  
 AUTHORS Isolation and characterization of the peptide synthetase genes phbB  
 TITLE and phcB from the phosphinothricin tripeptide producer Streptomyces  
 JOURNAL viridochromogenes  
 REFERENCE 2  
 AUTHORS Schwartz, D.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-MAY-1998) D. Schwartz, Universitaet Tuebingen,  
 Lehrstuhl Mikrobiologie/Biotechnologie, Eberhard-Karls-Universitaet  
 Tuebingen, Auf der Morgenstelle 28, D-72076 Tuebingen, FRG  
 revised by [3]  
 3 (bases 1 to 8992)  
 REFERENCE 3  
 AUTHORS Schwartz, D.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-AUG-2002) D. Schwartz, Universitaet Tuebingen,  
 Lehrstuhl Mikrobiologie/Biotechnologie, Eberhard-Karls-Universitaet  
 Tuebingen, Auf der Morgenstelle 28, D-72076 Tuebingen, FRG  
 On Aug 4, 2002 this sequence version replaced gi:8250615.  
 COMMENT Location/Qualifiers  
 FEATURES source  
 1..8992  
 /organism="Streptomyces viridochromogenes"  
 /mol\_type="genomic DNA"  
 /strain="Tue494"  
 /db\_xref="taxon:1938"  
 209..418  
 /note="ORF1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="CAB93682.1"  
 /db\_xref="GI:8250616"  
 /db\_xref="UniProt/TREMBL:Q9KMY9"  
 /translation="MIDGATREPLVYVNHGQYVSWPEDRPPAGKRRPHGTGRAC  
 LEVIRHMDRPLSLKSTFEEGA"  
 415..3984  
 /gene="phbB"  
 /gene="phbB"  
 415..3984  
 /gene="phbB"  
 /codon\_start=1  
 /transl\_table=11  
 /product="phosphinothricin tripeptide synthetase B, PhbB"  
 /protein\_id="CAB93683.2"  
 /db\_xref="GI:22095156"  
 /db\_xref="COA:Q9KMY8"  
 /db\_xref="UniProt/TREMBL:Q9KMY8"  
 /translation="MTAPQTDVVVGTGRIATWAGLLRPETIGIDNNVFRIGASSVAV  
 RAAARIRKALDTPLRLRDVFPSPALAKIRSRSTAPASPPTAPADSTATAP  
 LTFQBPWMLFDRMQPGNATYTIHFALRHGHLDTGLDRCVRDVVRRHVAFLRTSPS

VDRPQAOVUDRAHPIPGESDRLALPAAERPVAAARIAAEQAPELSTGPMIRVY  
 LHSIDTQRIULTWPHIVTDWSDDLIVRELNHLYRATHTGIAPOVHDMRA  
 ROBAELGTAEALLDWMHRLAGVPELLEPAPRPAAVRHRRGLLFIPESTVR  
 LEGIADKEGTTPYAVLLAGFAALLHRTIGODDLVVGSPAGRTTEGTVGLVNVY  
 AVRCDAKGRSPLEIVRTRTRVYSPAROEPLHRELAIPVRSAYPLVQVMA  
 LQNTPPDGRGRRGRRPARRESGDGCSANFDTLPTGSGACRGEWEDSDLE  
 RVASLGPQVTLDAALDRDLPALPIEDPDPARTAYLLRSGARVLLTLHRTNLPDADGT  
 PALLSAQHRTPDVALRTDGAELTYRQLHRLDRMVARRLSYLAPESVAVACSR  
 SPEWVVALLVAKAGCAVLPIDPDPAERTAYLLRSGARVLLTLHRTNLPDADGT  
 TVVTLDEPDSGMODTTSALPGIAPQOLAVLYLTSGSGRRPGVNLHGPVCRIRN  
 GMAFPFGPTIVLOKTPRIHPDVSVMEFWTATGATLVILAPDGHDPQVILGRLE  
 EGTVDHVPVSMALAFIDVICALPBGHSLRVRFGSGRLSPGLDRILPARIPIHEYNLH  
 YGTEALIEVTHMRCRGPETVPIGRITANARCTVLDALNPPVPGVGEMLAGVAV  
 ARGYHGRADLTARFLPDPYGPAGSRMYRSGDLARMRDVLLEYLGRDQVXLRGOR  
 LEIGELBATLAGAEVADVVDVVGCTGPODRIVAYVRPARGDBQLRTTLRELAAL  
 RLAPVRRSSVTLDRVPLTPSGTODKALPDPAAGOPRSGAAAGTGAERLGI  
 MRLIAGMGVGGDNPFEIGHSILAAHMTGRAPSTARGNDLVSLANAEHVLPAL  
 VTRQANTDSATERRLIAELADLTLEALPDDEGPDGSGE"  
 3977..5638  
 /note="ORF1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="CAB93276.1"  
 /db\_xref="GI:22095157"  
 /db\_xref="UniProt/TREMBL:Q9KJL4"  
 /translation="MSEETGAGSDTSDGTRGPRSGARPARADMAVAGVLIG  
 ALTCENILVIFPKTVAFGGGALVVVGAWMLRRLLSPSLPVVSGISGLSAI  
 AVSGIAANILKAGPPPVVALAVILANAVGLVZGMLAKRLSREILPPTLRPAI  
 LNMESLKRREVSARPLMSALSVVALVASVTDGTHALPGLPFGVGLASPMIL  
 VGVALAGSRSLAMWPFVGVGSAVFMAWRADPDYDTGHLGPAVALIGVITGVSVT  
 LVRSVSLRRSGALMWPVGDGAAACVLLAGALCMHGYAGRLGALLTWT  
 LFLALFEVRGAETGIALSPALFELGILRLMSDAIVSTVAGAMALVYT  
 LVAVSPSTAVSAPSPRIVSTOAGLAVGLFLMVAIVLSVAGGQEPPLAR  
 APDVASAYSLDSDAQOCTPLAALPAPFLITRAPVSSALGELLIPDTVLANS  
 LGGAAMLTVRKAPHLKETVGTVASGLVIGVSAIVARAFG"  
 5711..8971  
 /gene="phc"  
 /gene="phc"  
 5711..8971  
 /gene="phc"  
 /codon\_start=1  
 /transl\_table=11  
 /product="phosphinothricin tripeptide synthetase C, Phc"  
 /protein\_id="CAB93684.1"  
 /db\_xref="GI:8250618"  
 /db\_xref="COA:Q9KMY7"  
 /db\_xref="UniProt/TREMBL:Q9KMY7"  
 /translation="MEDLQTRIALSPKQALFESRLRAAARQPDALIRPDGDP  
 VPISFAQHRIMFLDQLEPGRPVNVNSLSLIGRPVTEAVRDLGALTTRHBYLRTVP  
 PADGSRHOIADSLTPELTETDLALPDSARAALRLCAEDKORFPDLSGFLAC  
 LLIRIRDDALLFTFHHVTFPDGMSIGLIRDLTALLHAETGDAGLPLPIQYADP  
 ADMQRRLDEKRLGELGYRERIRKAPPIYDIPDRPRPAAVTTBEARRRRPALP  
 TIALRDLAARSVAITPEFTMLTVPALHLHRSQSDWYIGIPVANARABELDILGFVA  
 NTLAMVRIRPEGMSFGDLAQVQVVEALAEATLELAQOMCDGSDVFGPTEATVTA  
 APFMEGDRDASELDLTLEPARDTHDPDSKRPDLTVLTREESTYGFPEYNALTFP  
 VTIDRLGERLALLARSVAADPGWELALPVLTDEIRHLEKVNAPVADDERHRTFG  
 VEDSARRHGHAVARADROLTYRLDEABANLAHLHLLIGRPRQPVVALDGTAD  
 AIVATVAVKAGAVLPIDPEYPAERLHILRSRGAULLTQSLAGRRAGNVYTYL  
 LDDATATLAALADGPADPGPIAIEDRIAYVITPSGSGVGVKGVNPRAGSLTRSSE  
 QFAQTPDSVLRFPASBSFVLSLEMLTPEFAGATVLEPPRALVPGEDLRLIRERYP  
 STVILSPSALSTLAGELPGLRTVVMAGAAATLELAQOMCDGSDVFGPTEATVTA  
 TIARCAPDRVPLGPVAGVAVVLDLDTLAPVFGSGELFLGCVGLARGVLDQPDVT  
 ADRFLDPDSGTEBERGARLYRTGDVVRWKAODELEFLRTHOVULRGRIILGELERT  
 EDHPGRTAVVIVRGESDRRLAGYVRAAGKRRPPTAGICRQLRDLPRYMWPELPL  
 VIALPTSPVNGKLDREALPDLAOSGDTGAGRPLDLPVERLSGIWQVVLGAPPS  
 ADNFPEVGSLSATRIIAVNOAFGRLPVRSLFVEPTLSGLARSASABREL"

ORIGIN  
 Query Match 78.3%; Score 18.8; DB 1; Length 8992;  
 Best Local Similarity 70.8%; Pred. No. 5.2e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCCGTGCGCGCGTSGACGCGSAC 24

Db 3420 GTCCTGCGGCGCGTCCCGGAC 3397

|||||:|||||:|||||

RESULT 20  
AE012334/C

LOCUS 11004 bp DNA linear BCT 23-MAY-2002  
DEFINITION Xanthomonas campestris pv. campestris str. ATCC 33913, section 242

ACCESSION AE012334 AE008922

VERSION AE012334.1 GI:21113402

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Xanthomonas campestris pv. campestris str. ATCC 33913  
Xanthomonas campestris pv. campestris str. ATCC 33913  
Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
Xanthomonadaceae; Xanthomonas.

1 (bases 1 to 11004)

da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,

Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida

Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,

Camargo, L.E.A., Camarotte, G., Camavari, F., Cardoso, J.,

Chamberg, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,

Cursino-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S.,

Ferreira, R.C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite

Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite

Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,

Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Melandri, J.,

Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,

Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.H.A.,

Rossi, A., Sana, J.A.D., Silva, C., de Souza, R.F., Spinoia, L.A.F.,

Takta, M.A., Tamura, R.E., Teixeira, E.C., Tezsa, R.I.D., Trindade dos

Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and

Kitejima, J.P.

Comparison of the genomes of two Xanthomonas pathogens with

differing host specificities

Nature 417 (6887), 459-463 (2002)

22022145

2 (bases 1 to 11004)

da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,

Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida

Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,

Camargo, L.E.A., Camarotte, G., Camavari, F., Cardoso, J.,

Chamberg, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,

Cursino-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S.,

Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,

Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite

Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,

Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Melandri, J.,

Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,

Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.H.A.,  
Rossi, A., Sana, J.A.D., Silva, C., de Souza, R.F., Spinoia, L.A.F.,  
Takta, M.A., Tamura, R.E., Teixeira, E.C., Tezsa, R.I.D., Trindade dos  
Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and  
Kitejima, J.P.  
Direct Submission  
Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de  
Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,  
Brazil  
FEATURES  
source  
1.11004  
Location/Qualifiers  
1.11004  
/organism="Xanthomonas campestris pv. campestris str. ATCC  
33913"  
/mol\_type="genomic DNA"  
/strain="ATCC 33913"  
/db\_xref="ATCC:33913"  
/db\_xref="taxon:190485"  
/note="pathovar: campestris"  
/gene="XCC2264"  
/complement(119. 607)  
/gene="XCC2264"  
/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer/genemark"

/codon\_start=1

/transl\_table=1

/product="general stress protein"

/protein\_id="AA041543.1"

/db\_xref="GI:21113403"

/translation="MADPKELQDKFWKALKSDRTVM/GLDGVDEGHARPMYAOIEGDS  
GGPIWFTTSKONALIAMLGGRRITGAFSSKGDLPISIGSLREDPDRAVYDRUNMP  
YVAAYVEGKDDPKLALLRDADHAQIWLNGSSILLAGIKVLLGVDPKKDQYQKADV  
LR"

1014. 2219

/gene="XCC2265"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer/genemark"

/codon\_start=1

/transl\_table=1

/product="pectinesterase"

/protein\_id="AA041544.1"

/db\_xref="GI:21113404"

/translation="MSASHRTTCLATALLSLIDNAALVTGTSRQLTSSNGTYT  
ISKLAAGAPITVALTDNMNPTAGVALTDFAADGSTRYRVAQADPAVAAGS  
TRRYSVKGATVELVCPGAPPLTFGSAATPDVIRYNNANPTPKPGTASHFC  
ASNSAATVGSNATYTVASGQARHLRPNNVAGCYADNNQSAVALAVRGDAS  
FEDVLVGNQDTLLISTNANVIRAFKSTTIEGDVDFIFSGITGFNDALITSAGA  
RLEGSRGQYIFAPTRGSGGFLAINSRVAGSGSDNDQTYGLARADVEGSLATV  
NGTSPNQVITRISTLSHRIKTPMNAVSTASBPFCSRCTNSANRFFSYATVALV  
PTDGRGVR"

2962. 4632

/gene="pgja"

/note="synonym: XCC2266"

2962. 4632

/gene="pgja"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer/genemark"

/codon\_start=1

/transl\_table=1

/product="polygalacturonase"

/protein\_id="AA041545.1"

/db\_xref="GI:21113405"

/translation="MOPHPRPALAPSLIALGLATPQASARPHGVADVTASIST  
SKGVQOPALPDSQCTTLKALVYVASSLADDDPQSRARLQAAIDVPCAGSA  
VRLVVGSGESGLSPPLTKSGVTLIDKVTLFGSRNPDQNGTCTATTDDA  
KSCNPLRHVADTASAVGAKIDGRGSLTVAGPNKASWMDLAVLTAKGSLDPA  
PRLQIDSTDTFTLIDTLKNSPHTTTNNVGLTAMGKITLAPSLVSRPGYCHA  
GSTDVNPBASCFPEPTAKNTDGFDPQSKVLLAYVIGTGDGVALIKAKAKRSI  
ASENMLFAYNPFYTHGSLGSESDSGMRHVAVGLSIDGNANDVQDPYSAGRLI  
KSDSRGQYDITFENICMGVARPIVPAVANYNPATROLPEFSGAILTNVSLGS  
KAFGGELSPFTGYDATTTLTSLDNVVDGGPVAPOHFGGSPNSGCAVHTFK  
GGPVSFPDLTBSANDVOLQGRPGAPALQCDDAFLAYHSLPDSPI"

complement(4695. 5384)

/gene="XCC2267"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer/genemark"

/codon\_start=1

/transl\_table=1

/product="conserved hypothetical protein"

/protein\_id="AA041546.1"

/db\_xref="GI:21113406"

/translation="MSQSALDSFLDKMRKMPMSVAEPFIPPORRLLAAMPALIQ  
EMEDIMNAGDPLPADAKLAWQOELDNWSQSRHPLGRVLEVPAPMQLQTLPA  
MOARAPDDIQAAMARLOAFBAVVAVFAVFAVTRSSSDATVAVVOMLDRORVAG  
SAAPGVGAPNRAQLRAMPKALARPVRVSRRLARLRRLAAGTAYAPRMQOL  
WHSWRASGAE"  
complement(5413. 6078)  
/gene="cbdz"  
/note="synonym: XCC2268"  
complement(5413. 6078)  
/gene="cbdz"  
/note="identified by sequence similarity; putative; ORF



located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=1  
/product="phosphoglycolate phosphatase"  
/protein\_id="AA041547.1"  
/db\_xref="GI:21133407"  
/translation="MRFRPRAVLPDLGLDLSADPMATUNAMISERGLPCTITLQGR  
PVKSGSRNALVAFAHLDAARALVPEFKRKEALGTQALPDGVWVLRLEQA  
GCWGIYTNKEBYLAOUILPQLGQORCAVIGDITLAEKRPPLPLVADRIQVAA  
TQCVYGDDESDIILAAARAAMPVAAALMGVRLGDDPLSMQADVAVQPPQLMBPAAW  
PQP"  
complement(6191. .6910)  
/gene="ubiG"  
/note="synonym: XCC2269"  
complement(6191. .6910)  
/gene="ubiG"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=1  
/product="3-demethylubiquinone-9-3-methyltransferase"  
/protein\_id="AA041548.1"  
/db\_xref="GI:21133408"  
/translation="MNPBPQSTSSNFHOTELDKTAALANRWMDADGQKPLHAPVR  
LDVVAARVALFGARVLDVCGGLSEMAALGAQVATIDAPLKVARLHLSSEV  
QVDYVQSVEDLAAEPQSPDPAVTCMELEHVPPLAIIRACASLIRKGLTFLSTLN  
RTPAFALAVAGAEYIARILPKGTHTYKDFIKPSELAAWLTAELOQDVSGMLYEPW  
RNRALSRTEVNTLACAVKP"  
complement(6939. .8288)  
/gene="XCC2270"  
complement(6939. .8288)  
/gene="XCC2270"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="AA041549.1"  
/db\_xref="GI:21133409"  
/translation="MDPLMTNAPPEPDLLEAGVVPVIEPFAVLEEDHVAVNSGV  
VAUPTDARVRPAPRTVSRPDAALMPLGLVNAHTNPMITLRVADDDHVAVNSGV  
IWPEAVITGPEFVADGTTLAIAMLRGTTVCNENYFPADVQAAYKKQHFRLVGA  
VVIDPTAMASDDEYFARAGELHDQWRDPLISTAFAPAPVTVNDANFERVMLAD  
QLDMPVHLHETAEQVADSVAGQORPLADLIGLVNDRLLIAVHTOLTEAHLIC  
AERGVSVHCPESNLKLAGFCFPCALORASVNIIGDGCASNNIDLMSEKNTAI  
LAKAVANDATLDAATTIRATLIGAGALRGCDIGISVEGKQNDYCVDSLAETSP  
LHVLSQILITAGSHQVTVNITAGKPKLVQRELIDMDPALVANARQWRDIRVRA"  
complement(8294. .9139)  
/gene="XCC2271"  
complement(8294. .9139)  
/gene="XCC2271"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=1  
/product="conserved hypothetical protein"  
/protein\_id="AA041550.1"  
/db\_xref="GI:21133410"  
/translation="MHLGRSGRAVRSASVLMVLSGLTAACPVEAQTVAACDAQVAA  
AVAAVVAAGVGAAGAAVAPARAAVAGTCKPVPYPSIRLAIAIAFAGDATTAGERDLE  
AVAAVVAAGVGAAGAAVAPARAAVAGTCKPVPYPSIRLAIAIAFAGDATTAGERDLE

Query Match 78.3%; Score 18.8; DB 1; Length 11004;  
Best Local Similarity 70.8%; Pred. No. 5.1e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCTGTGTGCGCGTGTACGCGSAC 24  
|||||:|||||:|||||:|||||:|||||:  
Db 1700 GTCTGTGTGCGCGTGTACGCGSAC 1677

RESULT 21  
AF506520  
LOCUS AF506520 13513 bp DNA linear BCT 16-DEC-2003

DEFINITION Streptomyces hygroscopicus AHBA gene cluster, partial sequence.  
ACCESSION AF506520  
VERSION AF506520.2 GI:39932993  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REMARK  
COMMENT  
FEATURES  
source  
1. 13513  
/organism="Streptomyces hygroscopicus"  
/mol\_type="genomic DNA"  
/strain="17997"  
/db\_xref="taxon:1912"  
1. .699  
/note="phosphatase"  
/codon\_start=1  
/transl\_table=1  
/product="Shnp"  
/protein\_id="AAG63810.1"  
/db\_xref="GI:24637549"  
/translation="MTRGPERFAAPTEIQRPPVPRHVAIFDLGVVDSPEWGEARS  
LAAVAVGCEAGPEEYRHRQGVYFDIMIMLPLEMBPPEYSYRLADRVQYNG  
VDVLTINRGRLAIATGKAGERSRLDVGILPYFAHVIGSDVPPKRPADII  
RRLLELVEAERAIMGDAPDILASAGADVAAVALMCCGAGAILAADPDVLRW  
PADILALCPALPGH"  
638. .1714  
/note="Shnk"  
/codon\_start=1  
/transl\_table=1  
/product="AHBA kinase"  
/protein\_id="AAR32671.1"  
/db\_xref="GI:39932994"  
/translation="MSGCGPPTCSPEARPCAPTEPPVAAAPSPWHGIDIGTKVAL  
RASADQRTERTFTLPENDDGDPDRDLALSDVAAALAAATGCGFOAVGAVATVCP  
GERVTSWPSRSPRTGLALPALRALVGAALIAADDDVAALEAAADANLLTYGV  
GTIGGGVILVGRSLCGPGRGSPFLGALVVALCGPGLTAAIAGSPATLAA  
RELHGEVTEYERLEALRTARPMATAIDTCALATAIGTLELGLAVIGCGFT  
AGIDNIDVRLATLHLETPSRPRPAPAPARIRLGLSLRGAVILAAVTVGAVPSVAA  
AAPRGDTRCAASTTPPALHQBRA"  
1715. .3202  
/note="Shn01"  
/codon\_start=1  
/transl\_table=1  
/product="cyclohex-1-ene-1-carboxylate CoA ligase"  
/protein\_id="AAR32672.1"  
/db\_xref="GI:39932995"  
/translation="MSGRYADQVEQFASLRLVGVGPGQVAVQLENNWETSALLIA  
VMTGAVVAPVTAALRELEKLAQLARVVTAVGVGWGYDVAGALAEAAERLPHIR  
HRVAVAGDAGGEMDFODHPORTPMAQOYMPDPEVDPOVALVETLACTGEPKV  
LHTSNMSSAAAGAVIGBERISGEASTPISLISITVLSVPSGSRATLFDKW  
APGRILAVIABARVSIITAAASRTIDIVRAVAAGNDTSRLTFSEGAIIPKLVRE  
VVELGALAKATYGTMLSTIGVTRPSDDPDMAAHSDGRGPAELDLRDAGVSKAN  
PAVLVYRAGACVAVVGLDGAVEILCDRDDGMLDGTGVAVPDGGRGIRTVGAVRI  
GDGFMIIPVDVGAQVLEHPAVAAVAVIAGRDGAVTVAAGPPTLEQLRTYL  
TAQGMAYVQPTLELVVPLPRNANGKVLKRELRGRVQSGSHA"





```

REFERENCE      Pseudomonadaceae; Pseudomonas.
AUTHORS        1 (bases 1 to 15356)
                Stover,C.K., Pham,X.-O.T., Ervin,A.L., Mizoguchi,S.D., Warrenner,P.,
                Hickey,M.J., Brinkman,F.S.U., Hutnagle,W.O., Kowalik,D.U.,
                Lagrou,M., Garber,R.L., Goltz,L., Tolentino,E.,
                Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
                Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
                Wong,G.K.-S., Wu,Z. and Paulsen,I.T.
TITLE          Complete genome sequence of Pseudomonas aeruginosa PA01, an
JOURNAL        opportunistic pathogen
MEDLINE        Nature 406 (6799), 959-964 (2000)
PUBMED         20437337
               10984043
REFERENCE      2 (bases 1 to 15356)
AUTHORS        Stover,C.K., Phan,X.-O.T., Ervin,A.L., Mizoguchi,S.D., Warrenner,P.,
                Hickey,M.J., Brinkman,F.S.U., Hutnagle,W.O., Kowalik,D.U.,
                Lagrou,M., Garber,R.L., Goltz,L., Tolentino,E.,
                Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
                Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
                Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,T., Sater,M.H.,
                Hancock,R.E.W., Lory,S. and Olson,M.V.
TITLE          Direct Submission
JOURNAL        Submitted (16-MAY-2000) Department of Medicine and Genetics,
               University of Washington Genome Center, University of Washington,
               Box 351145, Seattle, WA 98195, USA
               3 (bases 1 to 15356)
REFERENCE      . Pseudomonas aeruginosa Community Annotation Project (PseudocAP)
AUTHORS        Direct Submission
CONSRM         Submitted (04-FEB-2003) Department of Molecular Biology and
TITLE          Biochemistry, Simon Fraser University, 8888 University Dr.,
JOURNAL        Burnaby, British Columbia V5A 1S6, Canada
               -----
COMMENT        This represents the February 3, 2003 version of the continually
               updated, reviewed, Pseudomonas aeruginosa PAOI genome annotation,
               from PseudocAP (see http://www.pseudomonas.com for latest updates
               and links to alternate annotations). PseudocAP is coordinated by
               Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert
               E.W. Hancock (University of British Columbia, Canada). We welcome
               submission through www.pseudomonas.com of any proposed changes.
               -----
'Protein name confidence' is used to rate our confidence of the
accuracy of the protein name.
Class 1: Function experimentally demonstrated in P. aeruginosa.
Class 2: Function of highly similar gene experimentally
demonstrated in another organism (and gene context consistent
in terms of pathways its involved in, if known).
Class 3: Function proposed based on presence of conserved amino
acid motif, structural feature or limited sequence similarity
to an experimentally studied gene.
Class 4: Homology of previously reported genes of unknown function,
or no similarity to any previously reported sequences.
               -----
FEATURES       Location/Qualifiers
SOURCE         1..15356
                /organism="Pseudomonas aeruginosa PAOI"
                /mol_type="Genomic DNA"
                /strain="PAOI"
                /db_xref="taxon:208964"
                83..1207
                /gene="pilC"
                /locus_tag="PA4527"
                83..1207
                /gene="pilC"
                /locus_tag="PA4527"
                83..1207
                /gene="pilC"
                /locus_tag="PA4527"
                /note="Protein name confidence: Class 1 (function
                experimentally demonstrated in P. aeruginosa); Subcellular
                localization: Inner membrane protein; Subcellular
                localization confidence: Class 1; Alt protein name: pilin
                biogenesis protein PilC"
                /codon_start=1
                /transl_table=11
                /product="eF1L frameshift type 4 fimbrial biogenesis

```

```

gene
    protein Pf1C"
    /product="AAG07915.1"
    /db_xref="GI:9950770"
    /translation="MTLVKARLRKQGIPLAVRKKGISLDAGKKVKRMDIALPROMA
TMGAGVGLDPSFIIGEGPDNPNRKLTVDIEIKQEVSSGNSLANSLRKKQYIDELYC
NLVAVGEQSGALENLDRVATYKRTSLAKLRKANTYPIAVIVALIVSAILLIV
NPVAVGEQSGALENLDRVATYKRTSLAKLRKANTYPIAVIVALIVSAILLIV
VPOVAVGEQSGALENLDRVATYKRTSLAKLRKANTYPIAVIVALIVSAILLIV
DLTDDITLITLPIGGIYKKSAAVARTLTSTPAAGAPLVDALDSVGAAGNIVFKNA
VSKTKODVSTOMOLFNSWRTTSVFPNMAIQMTALGESSGSLDENLSVAASYEEVDN
AVDNLTLMEMPMIAVLGVLVAGLIVAMTLPIDLGNVG"
    /locus_tag="PA4528"
    /note="synonym: xcpa"
    1211..2083
    /gene="p1d"
    /locus_tag="PA4528"
    /note="Protein name confidence: Class 1 (function
experimentally demonstrated in P. aeruginosa); Subcellular
localization: Inner membrane protein; Subcellular
localization confidence: Class 1, Alt protein name: type 4
fimbrial biogenesis protein p1d"
    /codon_start=1
    /transl_table=1
    /product="type 4 prepilin peptidase p1d"
    /protein_id="AAG07916.1"
    /db_xref="GI:9950771"
    /translation="WPLIDYLAHPPLAFVLTCTILGLVGSPLNAVVHRLPRMERW
KAEAEALGLPERKQATYMLVLPNSACPRGHRIRWENIPVSYLATLGKSSCA
AIGKRYPLVEATLALSGVYAMHGFPTQACAMLLTMGLANSLIDADQLPDLVLA
PLMLGLIAHNGFLGSLDALFGAVGYSLTWSVWMLFVLYGKMGVGPDKLA
MIGANGGMOILLPTILSLVIGALGVIMLRNRAESGTPPEPFIATLNGVIALNG
DOITRTYIQPGFK"
    2080..2691
    /gene="coae"
    /locus_tag="PA4529"
    2080..2691
    /gene="coae"
    /locus_tag="PA4529"
    /note="Protein name confidence: Class 2 (high similarity
to functionally studied protein); Alt protein name: Coae
protein"
    /codon_start=1
    /transl_table=1
    /product="dephosphocoenzyme A kinase"
    /protein_id="AAG07917.1"
    /db_xref="GI:9950772"
    /translation="MTQPMILGLTGIGSGKSAEAHEPISGLVLDADNAAVNP
GRPALAKVERPGGILPDPQDLRALRERIPQAPEREPRMLDPLIGARTVOYVL
AAEASPAVLIVSPVSGQRQTHRETVLVDPTEHLOQTMLDCKSECVSSILDA
QARDERKRAADVVLVNDGDLHLQREVERLHAYTLIRGGA"
    2688..2888
    /locus_tag="PA4530"
    2688..2888
    /locus_tag="PA4530"
    /note="Protein name confidence: Class 4 (homologs of
previously reported genes of unknown function, or no
similarity to any previously reported sequences)"
    /codon_start=1
    /transl_table=1
    /product="conserved hypothetical protein"
    /protein_id="AAG07918.1"
    /db_xref="GI:9950773"
    /translation="MSQPLVTECPGCAPVWKSNDKRYRPFCSDRCLIDIGAAAE
HAIPDGLDDIDIFSADLPPEH"
    complement(2925..3134)
    /locus_tag="PA4531"
    complement(2925..3134)
    /locus_tag="PA4531"
    /note="Protein name confidence: Class 4 (homologs of
previously reported genes of unknown function, or no
similarity to any previously reported sequences)"
    /codon_start=1

```

/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG07919.1"  
/db\_xref="GI:9950774"  
/translation="WGHEHMAREGRELLSKSDEADRAMRLALLSAIVEINRAERDA  
GDIASELOFLAENLDDQRDYAFMRP"  
complement(3240..3929)  
/locus\_tag="PA4532"  
complement(3240..3929)  
/locus\_tag="PA4532"  
/note="Protein name confidence: Class 4 (homologs of  
previously reported genes of unknown function, or no  
similarity to any previously reported sequences)"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG07920.1"  
/db\_xref="GI:9950775"  
/translation="WTAABLSAESLEICMAIYAVVLLAAMRAFWELVSDTRQHL  
VEFTWLGFLMLVRDSESGISFPIGTATVTLDDPLAIWGLVAQLGLCLGRQ  
DMLAIVNGVLLVLAIPALIAELALFEKROPRLFYIFCGFPALITAVVLLG  
LGVWIDGLFPMFPWLEDPAQYLMIMPEAFINGTIVTGLVYVPEWLTFENRTRYL  
QAPKDERKRP"  
complement(3926..4396)  
/locus\_tag="PA4533"  
complement(3926..4396)  
/locus\_tag="PA4533"  
/note="Protein name confidence: Class 4 (homologs of  
previously reported genes of unknown function, or no  
similarity to any previously reported sequences)"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAG07921.1"  
/db\_xref="GI:9950776"  
/translation="MKRLHWVLLMLLGLAALAVKSGSTLEPWTLODPEKASINDS  
TRVLVARDMDSKLVKALAEKPGYLBARALTLADISMPALISRLFAPVAKRDY  
SYPVLIDRDAAIASRYASDEKVLWVRLQREVLTRQLDSPVALIALDQLP"  
complement(4393..4818)  
/locus\_tag="PA4534"  
complement(4393..4818)  
/locus\_tag="PA4534"  
/note="Protein name confidence: Class 4 (homologs of  
previously reported genes of unknown function, or no  
similarity to any previously reported sequences)"  
Query Match 78.3%; Score 18.8; DB 1; Length 15356;  
Best Local Similarity 70.8%; Pred.No.4.8e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
OY 1 GTCTGTGTGCGCGTACSGSAC 24  
|||||:||||:||||:||||:  
Db 11885 GTCTGTGTGCGCGCACCCGCAC 11862  
RESULT 23  
AY258138 34398 bp DNA linear BCT 02-JAN-2004  
LOCUS Pseudomonas aeruginosa strain C gene island PA01-4(C) sequence.  
DEFINITION  
ACCESSION AY258138  
VERSION AY258138.1 GI:33114121  
KEYWORDS  
SOURCE Pseudomonas aeruginosa  
ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
REFERENCE  
AUTHORS Klockgether J., Reva O.N., Larbig K. and Tummler B.  
TITLE Sequence Analysis of the Mobile Genome Island pKlC102 of  
Pseudomonas aeruginosa C  
JOURNAL J. Bacteriol. 186 (2), 518-534 (2004)  
PUBMED 14702321  
REFERENCE 2 (bases 1 to 34398)

AUTHORS Klockgether J., Reva O.N., Larbig K.D. and Tummler B.  
TITLE Direct Submission  
JOURNAL Submitted (19-MAR-2003) Klinische Forschungsgruppe, Medizinische  
Hochschule Hannover, Carl-Neuberg-Strasse 1, Hannover 30623,  
Germany  
FEATURES  
source  
location/Qualifiers  
1..34398  
/organism="Pseudomonas aeruginosa"  
/mol\_type="genomic DNA"  
/strain="C"  
/isolation\_source="airways of a cystic fibrosis patient"  
/db\_xref="taxon:287"  
/map="between genes PA0976 and PA1003"  
/clone="C"  
/note="hypervariable genome region"  
1..298  
/note="similar to Pseudomonas aeruginosa PA01 genomic  
region containing gene PA0976"  
223..298  
/product="tRNA-Lys"  
299..23693  
/note="gene island PA01-4(C)"  
complement(423..551)  
/gene="CL-1"  
complement(423..551)  
/gene="CL-1"  
/note="similar to PA0977; possibly truncated"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAP94679.1"  
/db\_xref="GI:33114122"  
/translation="MLAEPRSFKNTAVPSVIADYVEGALLRLQNGFRALETEKPT"  
complement(669..1949)  
/gene="xerc"  
/note="CL-2a"  
complement(669..1949)  
/gene="xerc"  
/note="phage-like integrase; similar to xerc-integrase  
from Pseudomonas aeruginosa plasmid pKlC102"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="putative integrase"  
/protein\_id="AAP94680.1"  
/db\_xref="GI:33114123"  
/translation="WTPQQLTEEYIPADHLREASAKIYRATKALNHPPTATVOEV  
DHRVILGRKRYLEQGLSKSWNTYSNLRITNGYATIEHLYTHSQVNPRTKTVIIP  
RRASKTYAAEXILIRRWMLNQVAGERTGDRAITPAWFLCTFEVYITGIRLNL  
LCIRKRDIDWNQILIRGETEKHKFEVPIITGLVPHLSRLAQEDRAFPADDDQL  
FNVRFSPPHKYSKMSNSDOVAMRTKLEKGVMTPIRFRHTLATDLMKAPENIHL  
TKCLNSNIQTTSYIEADYDHRVAVHARSLAQGLAENVRKYDVGSGSPASAKPKP  
CGQPLARWGVYPPPEARTPEAREPREHITPGTIOGGPYREBALQPPDTPQSVLFTL  
MAQHLNRAASASAAAPATSGSGGSGTANSSLA"  
1353..2042  
/gene="CL-2b"  
1353..2042  
/gene="CL-2b"  
/note="innerorf; similar to Pseudomonas aeruginosa  
plasmid pKlC102 ORF CP103b"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="putative excision regulator"  
/protein\_id="AAP94702.1"  
/db\_xref="GI:33114145"  
/translation="WRHKLRYMYHVELFMSFLCIAAEDQIIPVDVALADAQGRVOP  
DAGVEDIKRAKPPGSGDACAIGALAGADLHYEPILAGADGFCGDFACSPGDDDG  
GLSRVDLGVHOLVLDGIPPDROMRIYRPVAPROALFODLAPPSOHRPVVHLVG  
CRRAVLEQRIGRGAVDLRAGFSEIYVEDVILGELLRHSTIPASAVTVALGSRGC  
TGNSSLSGWSR"  
complement(1946..3772)  
gene

```

CDS
/gene="CL-3"
/complement(1946. .3772)
/note="similar to ORF CP102 from Pseudomonas aeruginosa
plasmid pKLC102"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAP94681.1"
/db_xref="GI:33114124"
/translation="MELIGTPRRRLLENTIMQASLSKQPEETIYRRLPIANVAYVOO
LPASENHHAHPCGMDHGLEIYVAYALKVRTYLLPTGGAESGSAQAEMASAAAG
ALANDIGKIVDLOVEILOSSTWHPMGPPIQYRFKTVRSRYQLGASALIHOL
LPRTALDWLSRFPBLMAQLYLFAQGYEHAIGLEIIVKADQASVQAEOLGPNPRLA
APKQSLORLADGLPLVKDFKFNQSGSDGWTODALMLVSKPADQCPAYLLAQ
GIEGSSNAPFPMLODAVITQNAEDKAIWTATINDGAGMRKFTLLKAPLIVA
DPAERPDYSGLIVVEGNSPEKPTTCEIIPNDPIORORPEAKMTLRDAPITLAP
SNDQIAKPSSTDQETDDYALGNTNSPPELDTSHSDSPANSPTANTRENTLOOP
LGTEPIDCAPEALIDVFMFSRTDLQGFVDMKSGITARLFINDKALVHTVDGT
AMLVTPGIKRYVQEHPEVEKLAQAKTAGKLVQRAFEKQGLHRTKTSKNINWTIVY
SGPKTKELKAYLLQDPKLLFPVQPLNPSLITVTDAGVE"
/complement(4196. .4546)
/gene="CL-4"
/complement(4196. .4546)
/note="similar to plasmid stabilization system protein
parB; similar to ORF CP93 from Pseudomonas aeruginosa
plasmid pKLC102"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAP94682.1"
/db_xref="GI:33114125"
/translation="MAKTRISHQAQDITIDLRTPHFGDPAARRVQALIGALEVY
ASDPQVGSISREELGAGLRSHLVYCRSMNVKAVRPRHFVFRVATQVLEVAV
LHDSMDLDQHLPHR"
/complement(4550. .4789)
/gene="CL-5"
/complement(4550. .4789)
/note="similar to predicted transcriptional regulators
containing the CopG/ArcMecJ DNA-binding domain (COG
3609); similar to ORF CP92 from Pseudomonas aeruginosa
plasmid pKLC102"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAP94683.1"
/db_xref="GI:33114126"
/translation="MEODINELVETGYQNRSEYIRAGRLILLQOEONSAKLEALN
ATSSGLAQLEGERDELSDLAQYLDDELQASH"
/complement(5311. .6822)
/gene="CL-6"
/complement(5311. .6822)
/note="similar to ORF CP91 from Pseudomonas aeruginosa
plasmid pKLC102"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAP94684.1"
/db_xref="GI:33114127"
/translation="MTFMTNDYLEYVYLLTGMIIINGIMNMSIDTGLFAVPAVAVR
EMLKVRGADENKGVLSARIEHTIYVGIYVALGIPVNVSPPTIEPDARAOO
COYVLPADTGSTSFSSLAGSKAPMAMMAHLSKETGAAVAATACGDTIRM
RMEYDNRVNNPLLAQEIADFSRCYCPSRRLFMKRPDGLSVEVDKALODLNLAS
RFLNLTGTYDIDYKSPROSQWPNATRDGLPVGQGGGIGPTCKQWMAOSGILWRS
IKQGVDDLMTSPLKMAKMLDQDEVTAVIRQVSPSGVKGAVTYTDTYGGQVGTWV
GIATAGTGVAGLSLAVFPAMDVVRQALPMVMSFLKMAVITCIPMLVIGTQLKVA

```

```

gene
/complement(6819. .7166)
/gene="CL-7"
/complement(6819. .7166)
/note="similar to ORF CP90 from Pseudomonas aeruginosa
plasmid pKLC102"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="hypothetical protein"
/protein_id="AAP94685.1"
/db_xref="GI:33114128"
/translation="MADTLRSRLQLQGLVGMIVVGLAVVGTLLSFLALNHRGIIQG
LEMPKSNVMSLPAWRALTYCALAIAMFRILKQKKEISAHRRQIRRIEIVLLILLI
EFKAVYRTGAA"
/complement(7166. .8548)
/gene="CL-8"
/complement(7166. .8548)
/note="similar to ORF CP89 from Pseudomonas aeruginosa
plasmid pKLC102"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAP94686.1"
/db_xref="GI:33114129"
/translation="MNITSVALTWLAAQLAQQDDPNNVSKTGTVLSDEVLYSIGGS
AVMSGSAQMDISIGVGMNNDLTCGMNISTYLENQNGATGSPQNIKSVIQNATG
AMSLPALIIRANPOLVNLITNGIILQARIDYDRSKGTCAIAEKADIAEGDTGK
IAGQALGATLASGDQAVSALFVEKKGNDVETWVGDKAGSSGQKPIRIIVDVTR
IAGNULTSRKVDSSVPSASGNNGVCTMSSPOBAAPATVGLGEOOQTCGCOOK
TYRAAGTLPRLIOETYDKLOSLOELSKRLETENLAACSTDLLPIRGVTEILR
DERDQVTLARLSVDVSLNDVSKALLLRLNFRAGAKEPVANAGLATQAVDQTSIL
QDEISINKLELELRRELASNPSPRVIERGQGRASGSSGVESAADADRLDRLQAPSA
GKSGSGGRP"

Query Match      78.3%; Score 18.8; DB 1; Length 34398;
Best Local Similarity 70.8%; Pred. No. 4.3e+03;
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Cy      1  GTCCGTGCGCCGTSACSGSAC 24
Db      18949  GTCCGCTGCGCGGTACCCCCAC 18972

RESULT 24
AF271693/c      50550 bp      DNA      linear      PHG 28-JAN-2001
LOCUS      AF271693
DEFINITION      Mycobacteriophage Bxb1, complete sequence.
ACCESSION      AF271693
VERSION      AF271693.1  GI:12583995
KEYWORDS
SOURCE
ORGANISM      Mycobacteriophage Bxb1
Virtues; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
REFERENCE
1 (bases 1 to 50550)
Medavilla,J., Jain,S., Kriakov,J., Ford,M.E., Duda,R.L.,
Jacobs,W.R., Hendrix,R.W. and Hatfull,G.F.
Genome organization and characterization of mycobacteriophage Bxb1
Mol. Microbiol. 38 (5), 955-970 (2000)
JOURNAL
MEDLINE      20572070
PUBMED      11123671
REFERENCE
2 (bases 1 to 50550)
Medavilla,J., Jain,S., Kriakov,J., Ford,M.E., Duda,R.L.,
Jacobs,W.R., Hendrix,R.W. and Hatfull,G.F.
Direct Submission
Submitted (24-MAY-2000) Biological Sciences, University of
Pittsburgh, Ruben Hall, Pittsburgh, PA 15260, USA
FEATURES
SOURCE
1. 50550

```

gene /organism="Mycobacteriophage Bxb1"  
CDS /mol\_type="genomic DNA"  
599..895 /db\_xref="taxon:148603"  
/gene="1"  
599..895  
/note="related to L5 gp4 (47%); 11.4 kD hypothetical  
protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp1"  
/protein\_id="AAG59706.1"  
/db\_xref="GI:12583996"  
/translation="MGWESSDRGRPLPADWTRIRREVLRAADHRCQIRYADICIGMAT  
EVDHYRDEESPLQASCKPCFKASMEGVAQRAKLAKMKRPPPRHGRSS"  
930..1370  
/gene="2"  
930..1370  
/note="related to L5 gp5 (41%); 16.3 kD hypothetical  
protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp2"  
/protein\_id="AAG59707.1"  
/db\_xref="GI:12583997"  
/translation="MGTGRGIPNRSDEVRNKEEYGVTTLPVSGPVKSPPLGLTDP  
HPVVDYINSLAESQALYQPSDHYAKFTLHPADOLKSKSPSQMLVAVNQMLSS  
LIVSGDRRRVTEIERTKSDGPDASVTTMGBLFRALKRKSS"  
1716..2027  
/gene="3"  
1716..2027  
/note="11.4 kD hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp3"  
/protein\_id="AAG59708.1"  
/db\_xref="GI:12583998"  
/translation="NAVVTYDNGQEWVHQHKYETDPDPOSPLILISNOGEVYGF  
NAQSWTHERVPMQIPVRIQDITLRKQSKKVAIVGDSYEPFGHITFSAEVAEKGLV  
v"  
2291..3361  
/gene="4"  
2291..3361  
/note="related to L5 gp31 (43%; 60aa); 36.7 kD  
hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp4"  
/protein\_id="AAG59709.1"  
/db\_xref="GI:12583999"  
/translation="MTVITLADVDRVTGQPNQQWVFSTVLRLQQDGSILITQKQVNP  
VUGALSVELPEGFALVYVGEYWFLEVPETDAGLWGLATSAVVPDTSAILLAQAVN  
GLIDANPPGADMLASNVSEPPSPASDVADVDLSAIAAYLASNPPEASVAMD  
DIDKSTPTSPSHSTSIANVTGLQDALDEKIDRAVAVARVSLGTAALVDASPTLDT  
LNELAALADDPRFATTVASQIGAKADKATTTTAGTGLGGCDLSANRTIYNSRSTSS  
TTACVGDNSRLSNTRIPTDGSVTNAAVAGAGIALSKLATGTVAGSDNSGANTLTITW  
GTEAOYTAIGTQDSNTIYLRTA"  
3371..4081  
/gene="5"  
3371..4081  
/note="related to L5 gp32 (31%; 60aa); 23.3 kD  
hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp5"  
/protein\_id="AAG59710.1"  
/db\_xref="GI:12584000"  
/translation="MAGMSLATTAFAKAAIGSTEIQKISIGTTEIWSAAPPTVDFDA

gene VSSFSNALGDLVSFSATAGARVFAVHLIGNNTVASTYGGNAMTLVTGAECITMNS  
CDS ASSGIVRLVTLASPGSGQTVENDRNANMAASPAISYLVASVGTSTPAAGTSTPS  
HSISAPPANGRVFQAFDMSSNVSLTSPGSGTGRINDRGTRAGMAARDSAAIYSTGTII  
SSTGAPVASIAIPMTPTV"  
4095..4355  
/gene="6"  
4095..4355  
/note="9.7 kD hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp6"  
/protein\_id="AAG59711.1"  
/db\_xref="GI:12584001"  
/translation="MTVTATVHDISGRPPDSHHTSSDLREQGVITTRVAVKPF  
NGEIALTLPPQPAVTHRQDRLIDVEEDSDMLDILLEATD"  
4363..4620  
/gene="7"  
4363..4620  
/note="related to TM4 gp90 (23%); 9.7 kD hypothetical  
protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp7"  
/protein\_id="AAG59712.1"  
/db\_xref="GI:12584002"  
/translation="TMNRLITMFAALVKAVEDYLRAHPEFLNQVIDRATAMPDLADI  
DDKIATKIPDLRDDKILIGMFPDLSRLPEQLINAINPFKA"  
4620..6053  
/gene="8"  
4620..6053  
/note="related to D29 gp10; H11415; 53.3 kD hypothetical  
protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp8"  
/protein\_id="AAG59713.1"  
/db\_xref="GI:12584003"  
/translation="MPRVVYGLTSSNGMPYLNDSGCEWTKIPGTSVTLQIAKQPLA  
ILRAPADPFAVYVLRDADASACWPTNNSVPSNHLSGTAMDLMNTHPEQVPAGFD  
AAKARVKEILDPIREGKVPFNGNDSSPRDANHPLAISLRNGTDPDTGNPTADPIAR  
KIRADGYSTFERGSAPASAPITLAAATLSARAAETLIPAVRSGLRBSECTNVRIAM  
WIAOIGHESGSFOYTEIEIAKNRVAPIIGRTWIOITWIDYNRSFSQAVARQVPTPD  
YFVNVYRELADLKWAGIGPAMYVTVARPDINELDRDLNVTTRIRINGTGLADROA  
RYNRALQGDALLIOLHEDEDFLSALTDAEORELIDLARQOAKYKRSRSPLHWPEG  
EVDITAGSMTGDNVHIQIVYEKLAIVTIGDVSIALIAYVNSDSDPTNPELAKRIK  
RVKPEDITAAQVOIQKWLAAEQFHAA"  
6043..7011  
/gene="9"  
6043..7011  
/note="related to L5 gp12 (38%); 35.9 kD hypothetical  
protein"  
/codon\_start=1  
/transl\_table=11  
/product="gp9"  
/protein\_id="AAG59714.1"  
/db\_xref="GI:12584004"  
/translation="WPLKLGDRNPTVYRWKREVMARFAGYARVHGPIPTDDEGRPA  
EAMQVEYSRTFQPLDGIIVSDDLRSIGIPAEEDARPVLLVSGTGVFWVGPDAVA  
RRLGQVYLMRPVGPYTAQAFPMGSPVANVTETATRLIEERRRLERGLSMIGSOG  
AIVTSELMWEYHIKPYTGRLHWKDHVRGAATFGNPMRETKVMPDGGQMSAKSHGI  
ADQLMVDTPDMWRVYAHKGDLYTDCGSGSGEMKTAIVKVMMSRVFSGSDSLRQLLE  
IGVNPTEFLILIRAVLDAGLFFTRIGTGPHVNYINDPATDRLRSYT"  
7031..8719  
/gene="10"  
7031..8719  
/note="related to L5 gp13 (65%); 62.3 kD putative  
terminase protein"  
/codon\_start=1



\* 1158 1257: gap of unknown length  
\* 1258 2321: contig of 1064 bp in length  
\* 2322 2421: gap of unknown length  
\* 2422 3618: contig of 1197 bp in length  
\* 3619 3718: gap of unknown length  
\* 3719 4930: contig of 1212 bp in length  
\* 4931 5030: gap of unknown length  
\* 5031 6031: contig of 1001 bp in length  
\* 6032 6131: gap of unknown length  
\* 6132 7270: contig of 1139 bp in length  
\* 7271 7370: gap of unknown length  
\* 7371 8508: contig of 1138 bp in length  
\* 8509 8608: gap of unknown length  
\* 8609 10588: gap of 1980 bp in length  
\* 10589 10688: gap of unknown length  
\* 10689 12189: contig of 1501 bp in length  
\* 12190 12289: gap of unknown length  
\* 12290 14285: contig of 1996 bp in length  
\* 14286 14385: gap of unknown length  
\* 14386 15907: contig of 1522 bp in length  
\* 15908 16007: gap of unknown length  
\* 16008 17891: contig of 1884 bp in length  
\* 17892 17991: gap of unknown length  
\* 17992 19622: contig of 1631 bp in length  
\* 19623 19722: gap of unknown length  
\* 19723 21796: contig of 2074 bp in length  
\* 21797 21896: gap of unknown length  
\* 21897 24606: contig of 2710 bp in length  
\* 24607 24706: gap of unknown length  
\* 24707 26618: contig of 1912 bp in length  
\* 26619 26718: gap of unknown length  
\* 26719 28720: contig of 2002 bp in length  
\* 28721 28820: gap of unknown length  
\* 28821 31309: contig of 2489 bp in length  
\* 31310 31409: gap of unknown length  
\* 31410 34084: contig of 2675 bp in length  
\* 34085 34184: gap of unknown length  
\* 34185 36260: contig of 2076 bp in length  
\* 36261 36360: gap of unknown length  
\* 36361 37424: contig of 1064 bp in length  
\* 37425 37524: gap of unknown length  
\* 37525 39771: contig of 2247 bp in length  
\* 39772 39871: gap of unknown length  
\* 39872 43194: contig of 3323 bp in length  
\* 43195 43294: gap of unknown length  
\* 43295 46776: contig of 3482 bp in length  
\* 46777 46876: gap of unknown length  
\* 46877 51203: contig of 4327 bp in length  
\* 51204 51303: gap of unknown length  
\* 51304 53821: contig of 2518 bp in length  
\* 53822 53921: gap of unknown length  
\* 53922 56371: contig of 2450 bp in length  
\* 56372 56471: gap of unknown length  
\* 56472 60412: contig of 3941 bp in length  
\* 60413 60512: gap of unknown length  
\* 60513 65444: contig of 4932 bp in length  
\* 65445 65544: gap of unknown length  
\* 65545 71808: contig of 6264 bp in length  
\* 71809 71908: gap of unknown length  
\* 71909 81306: contig of 9388 bp in length  
\* 81307 81406: gap of unknown length  
\* 81407 94058: contig of 12652 bp in length.

FEATURES  
source

1..94058  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-482H23"

## ORIGIN

Query Match 78.3%; Score 18.8; DB 2; Length 94058;  
Best Local Similarity 70.8%; Pred. No. 3.7e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Cy 1 GTCTGSGTGGCCGTAACGCGAC 24  
Db 9282 GTCTGGGGGCGCGAGCGCAC 9259

RESULT 26  
AP006618\_13  
WPCOMMENT  
Sequence split into 61 fragments LOCUS AP006618 Accession AP006618

Fragment Name	Begin	End
AP006618_00	1	110000
AP006618_01	100001	210000
AP006618_02	200001	310000
AP006618_03	300001	410000
AP006618_04	400001	510000
AP006618_05	500001	610000
AP006618_06	600001	710000
AP006618_07	700001	810000
AP006618_08	800001	910000
AP006618_09	900001	1010000
AP006618_10	1000001	1110000
AP006618_11	1100001	1210000
AP006618_12	1200001	1310000
AP006618_13	1300001	1410000
AP006618_14	1400001	1510000
AP006618_15	1500001	1610000
AP006618_16	1600001	1710000
AP006618_17	1700001	1810000
AP006618_18	1800001	1910000
AP006618_19	1900001	2010000
AP006618_20	2000001	2110000
AP006618_21	2100001	2210000
AP006618_22	2200001	2310000
AP006618_23	2300001	2410000
AP006618_24	2400001	2510000
AP006618_25	2500001	2610000
AP006618_26	2600001	2710000
AP006618_27	2700001	2810000
AP006618_28	2800001	2910000
AP006618_29	2900001	3010000
AP006618_30	3000001	3110000
AP006618_31	3100001	3210000
AP006618_32	3200001	3310000
AP006618_33	3300001	3410000
AP006618_34	3400001	3510000
AP006618_35	3500001	3610000
AP006618_36	3600001	3710000
AP006618_37	3700001	3810000
AP006618_38	3800001	3910000
AP006618_39	3900001	4010000
AP006618_40	4000001	4110000
AP006618_41	4100001	4210000
AP006618_42	4200001	4310000
AP006618_43	4300001	4410000
AP006618_44	4400001	4510000
AP006618_45	4500001	4610000
AP006618_46	4600001	4710000
AP006618_47	4700001	4810000
AP006618_48	4800001	4910000
AP006618_49	4900001	5010000
AP006618_50	5000001	5110000
AP006618_51	5100001	5210000
AP006618_52	5200001	5310000
AP006618_53	5300001	5410000
AP006618_54	5400001	5510000
AP006618_55	5500001	5610000
AP006618_56	5600001	5710000
AP006618_57	5700001	5810000
AP006618_58	5800001	5910000
AP006618_59	5900001	6010000
AP006618_60	6000001	6021225

Continuation (14 of 61) of AP006618 from base 1300001 (AP006618 Nocardia farcinica ITM 1

Query Match 78.3%; Score 18.8; DB 1; Length 110000;  
 Best Local Similarity 70.8%; Pred. No. 3.6e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTCCGTGCGCCGCGACGCGAC 24  
 Db 86535 GTCCGTGCGCCGCGACGCGAC 86558

RESULT 27  
 CN508CAM 146585 bp DNA linear PLN 21-NOV-2003  
 LOCUS Oryza sativa chromosome 12, BAC OJ1268\_D02 of library Monsanto  
 DEFINITION from chromosome 12 of cultivar Nipponbare of ssp. japonica of Oryza  
 sativa (rice), complete sequence.

ACCESSION AL831809 GI:25900557  
 VERSION HTG  
 KEYWORDS Oryza sativa (japonica cultivar-group)  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 146585)  
 AUTHORS Chichima, N., Orjeda, G., Catolico, L., Demange, N., Wincker, P.,  
 Segurens, B., Pelletier, E., Scarpelli, C., Salanoubat, M.,  
 Weissenbach, J. and Quelier, F.

TITLE Oryza sativa chromosome 12 sequencing  
 JOURNAL Unpublished  
 AUTHORS 2 (bases 1 to 146585)  
 REFERENCE GenomeScope.

COMMENT Direct Submission  
 Submitted (20-NOV-2003) GenomeScope - Centre National de Sequencage :  
 BP 191 91006 Evry cedex - FRANCE (E-mail : segref@genomescope.cns.fr  
 On Nov 28, 2002 this sequence version replaced gi:24817645.  
 Center: GenomeScope / Centre National de Sequencage  
 Center code: GS  
 Web site: http://www.genomescope.cns.fr/  
 Contact: Segref@genomescope.cns.fr

-----  
 The following sequence is oriented from the T7 to the SP6 end. The  
 nucleotide sequence of this BAC clone was generated by combining  
 Monsanto and GenomeScope sequencing data.  
 Upstream BAC (overlapping the T7 end) : OJ1559\_C07 (AC=AL731187)  
 Downstream BAC (overlapping the SP6 end) : OJ1119\_E02 (AC=AL731762)  
 -----  
 FINISHED SEGMENT STARTS AT BASE 1  
 FINISHED SEGMENT ENDS AT BASE 146585  
 -----

FEATURES  
 Source location/Qualifiers  
 1..146585  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /cultivar="Nipponbare"  
 /sub\_species="japonica"  
 /db\_xref="taxon:39947"  
 /chromosome="12"  
 /clone="OJ1268\_D02"  
 /clone\_1db="Monsanto"

ORIGIN  
 Query Match 78.3%; Score 18.8; DB 8; Length 146585;  
 Best Local Similarity 70.8%; Pred. No. 3.5e+03;  
 Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTCCGTGCGCCGCGACGCGAC 24  
 Db 102468 GTCCGTGCGCCGCGACGCGAC 102491

RESULT 28

AC084284/c  
 LOCUS AC084284 152186 bp DNA linear HTG 11-MAR-2001  
 DEFINITION Homo sapiens chromosome RPL1-11 clone RPL1-276D3, WORKING DRAFT  
 SEQUENCE, 30 unordered pieces.  
 AC084284  
 ACCESSION AC084284.3 GI:13270752  
 VERSION HTG  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 152186)  
 Waterston, R.H.  
 The sequence of Homo sapiens clone  
 Unpublished  
 2 (bases 1 to 152186)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (20-OCT-2000) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 On Mar 10, 2001 this sequence version replaced gi:10945792.

-----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 Project Information  
 Center project name: H\_NH0276D03  
 Summary Statistics  
 Sequencing vector: plasmid, 0%  
 Chemistry: Dye-terminator Big Dye; 0% of reads  
 Assembly program: Phrap; version 0.99019  
 Consensus quality: 132621 bases at least Q40  
 Consensus quality: 139102 bases at least Q30  
 Consensus quality: 142309 bases at least Q20  
 Insert size: 158000; agarose-fp  
 Insert size: 149286; sum-of-coverage  
 Quality coverage: 3.13 in Q20 bases; agarose-fp  
 Quality coverage: 3.40 in Q20 bases; sum-of-coverage  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 30 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1  
 1511: contig of 1511 bp in length  
 1512  
 1611: gap of unknown length  
 1612  
 2942: contig of 1331 bp in length  
 2943  
 3042: gap of unknown length  
 3043  
 4265: contig of 1223 bp in length  
 4266  
 4365: gap of unknown length  
 4366  
 6324: contig of 1959 bp in length  
 6325  
 6424: gap of unknown length  
 6425  
 7995: contig of 1571 bp in length  
 7996  
 8095: gap of unknown length  
 8096  
 9575: contig of 1480 bp in length  
 9576  
 9675: gap of unknown length  
 9676  
 11960: contig of 2285 bp in length  
 11961  
 12060: gap of unknown length  
 12061  
 14652: contig of 2592 bp in length  
 14653  
 14752: gap of unknown length  
 14753  
 17032: contig of 2280 bp in length  
 17033  
 17132: gap of unknown length  
 17133  
 19636: contig of 2504 bp in length  
 19637  
 19736: gap of unknown length  
 19737  
 23580: contig of 3844 bp in length  
 23581  
 23680: gap of unknown length



```

* 23681 26611: contig of 2931 bp in length
* 26612 26711: gap of unknown length
* 26712 29432: contig of 2721 bp in length
* 29433 29532: gap of unknown length
* 29533 34556: contig of 5024 bp in length
* 34557 34656: gap of unknown length
* 34657 37667: contig of 2911 bp in length
* 37668 41750: contig of 4083 bp in length
* 41751 47006: contig of 5156 bp in length
* 47007 52437: contig of 5331 bp in length
* 52438 52537: gap of unknown length
* 52538 57297: contig of 4760 bp in length
* 57298 62529: contig of 5132 bp in length
* 62530 62629: gap of unknown length
* 62630 66478: contig of 3849 bp in length
* 66479 75759: contig of 9181 bp in length
* 75760 75859: gap of unknown length
* 75860 83433: contig of 7574 bp in length
* 83434 83533: gap of unknown length
* 83534 93430: contig of 9897 bp in length
* 93431 100839: contig of 7309 bp in length
* 100840 100939: gap of unknown length
* 100940 109890: contig of 8951 bp in length
* 109891 109990: gap of unknown length
* 109991 119771: contig of 9781 bp in length
* 119772 119871: gap of unknown length
* 119872 128562: contig of 8691 bp in length
* 128563 140541: contig of 11879 bp in length
* 140542 152186: contig of 11545 bp in length.
* 140642

```

```

FEATURES
source
1..152186
Location/Qualifiers

```

```

misc_feature /note="assembly_name:Contig11"
misc_feature /note="assembly_name:Contig12"
misc_feature /note="assembly_name:Contig13"
misc_feature /note="assembly_name:Contig14"
misc_feature /note="assembly_name:Contig15"
misc_feature /note="assembly_name:Contig16"
misc_feature /note="assembly_name:Contig17"
misc_feature /note="assembly_name:Contig18"
misc_feature /note="assembly_name:Contig19"
misc_feature /note="assembly_name:Contig20"
misc_feature /note="assembly_name:Contig21"
misc_feature /note="assembly_name:Contig22"
misc_feature /note="assembly_name:Contig23"
misc_feature /note="assembly_name:Contig24"
misc_feature

```

```

misc_feature /note="assembly_name:Contig25"
misc_feature /note="assembly_name:Contig26"
misc_feature /note="assembly_name:Contig27"
misc_feature /note="assembly_name:Contig28"
misc_feature /note="assembly_name:Contig29"
misc_feature /note="assembly_name:Contig30"
misc_feature /note="assembly_name:Contig31"
misc_feature /note="assembly_name:Contig32"
misc_feature /note="assembly_name:Contig33"
misc_feature /note="assembly_name:Contig34"
misc_feature /note="assembly_name:Contig35"
misc_feature /note="assembly_name:Contig36"
misc_feature /note="assembly_name:Contig37"
misc_feature /note="assembly_name:Contig38"
misc_feature /note="assembly_name:Contig39"
misc_feature /note="assembly_name:Contig40"

```

# ORIGIN

```

Query Match 78.3%; Score 18.8; DB 2; Length 152186;
Best Local Similarity 70.8%; Pred. No. 3.4e+07;
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 GTCCTGCTGCGCGTACGCGSAC 24
DB 95430 GACCTGCTGCGCGTACGCGCAC 95407

```

```

RESULT 29
AL663032 162995 bp DNA linear ROD 05-APR-2002
LOCUS Mouse DNA sequence from clone RP23-2716 on chromosome X, complete
DEFINITION
ACCESSION AL663032
VERSION AL663032.9 GI:20068678
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Whitehead,S.
TITLE Direct Submision
JOURNAL Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
hunquerry@sanger.ac.uk Clone request: clonerequest@sanger.ac.uk
On Apr 7, 2002 this sequence version replaced gi:1935849.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the

```



assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-2716 is from the RPECI-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/dacpac/home.htm> VECTOR: pBAC3.6.

# FEATURES

## Source

1.162995  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="X"  
/clone="RP23-2716"  
/clone\_id="RPECI-23"

## ORIGIN

Query Match 78.3%; Score 18.8; DB 10; Length 162995;  
Best Local Similarity 70.8%; Pred. No. 3.4e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTCCTGTCGCGCGTACCGGAC 24

Db 158767 GTCCTGGGGGCGCGGACCGGCAC 158790

## RESULT 30

AL646062

LOCUS Ralstonia solanacearum GM11000 chromosome, complete sequence;  
segment 6/19

ACCESSION AL646062  
VERSION AL646062.1 GI:17427974

KEYWORDS

SOURCE

ORGANISM Ralstonia solanacearum

Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; Ralstonia.

## REFERENCE

AUTHORS

1 Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S., Ariat,M., Billault,A., Broctier,P., Camus,J.C., Catolico,L., Chandier,M., Choisme,N., Claudel-Renard,C., Cunnac,S., Demange,N., Gaepin,C., Lavie,M., Moisan,A., Robert,C., Saurin,M., Schiek,T., Signier,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M., Weissenbach,J. and Boucher,C.A.  
Genome sequence of the plant pathogen Ralstonia solanacearum Nature 415 (6871), 497-502 (2002)

TITLE

JOURNAL

MEDLINE

PUBMED

2 (bases 1 to 193050)

REFERENCE

Boucher,C.A.

TITLE

JOURNAL

Submitted (05-DEC-2001) GenomeScope and CNRS UMR-8030, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS, BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMM CNRS 118 Route de Narbonne, F 31062 Toulouse Cedex 4, GenomeScope and INRA URV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, F31326 Castanet-Tolosan Cedex, Laboratoire de Génétique Cellulaire INRA, BP27, F31326 Castanet-Tolosan Cedex, Laboratoire de Génétique Cellulaire Christian.Boucher@toulouse.inra.fr  
<http://sequence.toulouse.inra.fr/R.solanacearum.html>.

FEATURES

Source

1.193050  
/organism="Ralstonia solanacearum"  
/mol\_type="genomic DNA"  
/strain="GM11000"  
/db\_xref="taxon:305"

gene

CDS

gene

CDS

CDS

505..771  
/gene="RSC0962"  
/note="synonym: RS04402"  
505..771  
/gene="RSC0962"  
/functions="elements of external origin; phage-related functions and prophages"  
/note="Product confidence : putative Gene name confidence : hypothetical predicted by Codon usage predicted by Homology predicted by Framed"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="PUTATIVE BACTERIOPHAGE DNA-BINDING TRANSCRIPTION REGULATOR PROTEIN"  
/protein\_id="CAD14664.1"  
/db\_xref="GI:17427975"  
/db\_xref="UniProt/TREMBL:O8Y0T1"  
/translation="MKYORAHAVAAACSTPAPAVSEATARFLRLPEVMSACGUS RSSYGVSGGEPFQPVRLPSSVAVWDSIRQWIAARIAROS"  
789..1517  
/gene="RSC0963"  
/note="synonym: RS04401"  
789..1517  
/gene="RSC0963"  
/functions="miscellaneous; unknown"  
/note="Product confidence : hypothetical Gene name confidence : hypothetical predicted by Codon usage predicted by Framed"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="HYPOTHEITICAL PROTEIN"  
/protein\_id="CAD14665.1"  
/db\_xref="GI:17427976"  
/db\_xref="UniProt/TREMBL:O8Y0T0"  
/translation="MKONTVALPENTLOALTRHONARGRFPAARDEADRIADMOX HRTNADSETPAQARQEAQKMRSTTAQDRLDKAKRAAYATREDYRAIYAE EAHVADAKIGAVAKLDENAAVRYLVQYADTLMSAKOMLVPLRAIAVERAYTSGA TRDIAAEVYNESADALARMYGVIKQAFARFPAAGVAVLOAIRPADLRFIP SPQAIHVRNTLRKAAQQRGRAA"  
1811..2032  
/gene="RSC0964"  
/note="synonym: RS04403"  
1811..2032  
/gene="RSC0964"  
/functions="elements of external origin; phage-related functions and prophages"  
/note="Product confidence : putative Gene name confidence : hypothetical predicted by Codon usage predicted by Homology predicted by Framed"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="PUTATIVE BACTERIOPHAGE TRANSCRIPTIONAL ACTIVATOR-RELATED TRANSCRIPTION REGULATOR PROTEIN"  
/protein\_id="CAD14666.1"  
/db\_xref="GI:17427977"  
/db\_xref="GOA:Q810S9"  
/db\_xref="UniProt/TREMBL:O8Y0S9"  
/translation="MNCPCGSAVATRTSPVSRITRELYCQCSNVLGHTFVSLVEV VRTLSPTSDPVPARQLAGRSEPVAAV"  
2400..2576  
/gene="RSC0965"  
/note="synonym: RS04404"  
2400..2576  
/gene="RSC0965"  
/functions="miscellaneous; unknown"

/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="HYPOTHETICAL PROTEIN"  
/protein\_id="CAD14667.1"  
/db\_xref="GI:17427978"  
/db\_xref="UniProt/TREMBL:O8Y0S8"  
/translation="MLDAPARPEQPAFPQIALIVTALRDVAAPTDAASLDVAGAL  
LAAVAIAQARRRHG"  
/gene="RSC0966"  
/note="synonym: RS04400"  
2569. 2892  
/gene="RSC0966"  
/function="miscellaneous; hypothetical/global homology"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="CONSERVED HYPOTHETICAL PROTEIN"  
/protein\_id="CAD14668.1"  
/db\_xref="GI:17427979"  
/db\_xref="UniProt/TREMBL:O8Y0S7"  
/translation="MKRNHPHWALEKQSLTQLAQOAPQTPPAATLARLETVS  
ASMRDATERAGQLCVSALDGLCLDLHAQTAPGYRSILCLLVKQQLDGAIGT  
VTBML"  
2903. 5704  
/gene="RSC0967"  
/note="synonym: RS04399"  
2903. 5704  
/gene="RSC0967"  
/function="miscellaneous; unknown"  
/note="Product confidence : hypothetical  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="HYPOTHETICAL PROTEIN"  
/protein\_id="CAD14669.1"  
/db\_xref="GI:17427980"  
/db\_xref="GOA:O8Y0S6"  
/db\_xref="UniProt/TREMBL:O8Y0S6"  
/translation="MHTQSPALRTWTFDQDKRIDLDHLAERDLRRSGAKGNYS  
HTRDSASTSIFPERGAKMDSTEGSSCIDYOHCMQDVATPMEAKILGQWYGI  
PMP  
AAPASALLARKSTERYIARRQANPEPAAVYLAGGVDEAVSRRALQTGALGNMNS  
PVPAGEAGHGPPAAFIVRANDTAVVADLRVADPAINGVNTQCGDKLGHMTS  
DARRLHAHTVIVESPINALVECHLPEGTAAALRTANVDKIWTFLKRGVVI  
ALDHTDPVNERGTMRGLAANKLSEALTAADIGAMLVMDQMEGEEDINDVLAHGA  
AELTARMRLKAMLIKGMFGGERLSTGRVFLPHDFGIYRFRVKEDFTQYVEFK  
DADDEGOTSRSEBELGICAPRVAGLSRLRVSHLATINGTADSOPEYFEGTSAOVAR  
HGATLQREYVGNDKLYNLEWFKGFGHTMPESOPARMNVYLERADDLGRDYVNVGL  
AWRGELALLESDCYVEPQOCLYNNRFPRTSQSRAVIDAVQATFKGNAATA  
LVWALGHLKTVLGFPYHLOMAEKAGKAGKLESIOATASQVLSGQMLKTDHRRRA  
SVSWTSHPVGMDSPSKLPAVLSIEDGLQSTYREFETVGAALTPYLMCAVLLAGE  
EVDVESLQSKICRTSLSDVKQAILPRNLPOPVNAWLOFLASQOBERIDLAAAFD  
VLSRSRATNDATAPRWENYAAITWATAALAEPAQIDVEQGGPDIHILVEMNAHIA  
ETDGTRLPVWYIMEILSELDAGRFEPHCMTPTIQTDSGREMALFLRSHVMDHSTA  
MHLRGKFDALPIKTARVFKQQLQSGVVPQHGKLEDEVERKILGRCAHMAAISLPR  
LEORGLYASPALAAAYOMAS"  
complement (6298. 7473)  
/gene="RSC0968"  
/note="synonym: RS04398"  
complement (6298. 7473)

/gene="RSC0968"  
/functions="elements of external origin; transposon-related  
functions"  
/note="Product confidence : probable  
Gene name confidence : hypothetical  
predicted by Codon usage  
predicted by Homology  
predicted by Framed"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="PROBABLE CP4-LIKE INTEGRASE PROTEIN"  
Query Match 78.3%; Score 18.8; DB 1; Length 193050;  
Best Local Similarity 70.8%; Pred. No. 3.3e+03;  
Matches 17; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GTCCGTGCGCGGTGACGCGSAC 24  
Db 78885 GCCCTGCTGCGCGGTGACGCGCAC 78908  
Search completed: July 20, 2005, 16:45:50  
Job time : 1178.57 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2005, 16:46:04 ; Search time 429.857 Seconds

(without alignments)  
265.662 Million cells updates/sec

Title: US-10-017-471b-1

Perfect score: 18

Sequence: 1 gaccacgctccggagcatg 18

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 7173243 seqs, 3172129809 residue

Total number of hits satisfying chosen parameters: 14346486

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*  
19: /cgn2\_6/ptodata/2/pubpna/US10G\_PUBCOMB.seq:\*  
20: /cgn2\_6/ptodata/2/pubpna/US10H\_PUBCOMB.seq:\*  
21: /cgn2\_6/ptodata/2/pubpna/US10I\_PUBCOMB.seq:\*  
22: /cgn2\_6/ptodata/2/pubpna/US10J\_NEW\_PUB.seq:\*  
23: /cgn2\_6/ptodata/2/pubpna/US11A\_PUBCOMB.seq:\*  
24: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq:\*  
25: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
26: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17.2	95.6	18	15	US-10-017-471a-1
2	17.2	95.6	891	15	US-10-156-761-3511
3	17.2	95.6	9025608	15	US-10-156-761-1
4	16	88.9	4346	15	US-10-017-471a-19
5	15.6	86.7	513	10	US-09-975-719-155
6	15.6	86.7	612	22	US-10-617-320-417
7	15.6	86.7	762	10	US-09-975-719-153

C 8	15.6	86.7	1080	13	US-10-027-632-100941	Sequence 100941,
C 9	15.6	86.7	1080	13	US-10-027-632-100942	Sequence 100942,
C 10	15.6	86.7	1080	13	US-10-027-632-100943	Sequence 100943,
C 11	15.6	86.7	1080	17	US-10-027-632-100941	Sequence 100941,
C 12	15.6	86.7	1080	17	US-10-027-632-100942	Sequence 100942,
C 13	15.6	86.7	1080	17	US-10-027-632-100943	Sequence 100943,
C 14	15.6	86.7	1080	21	US-10-472-928-473	Sequence 473, App
C 15	15.6	86.7	2048	10	US-09-975-719-136	Sequence 136, App
C 16	15.6	86.7	2395	8	US-08-961-527-205	Sequence 205, App
C 17	15.6	86.7	2395	13	US-10-158-844-205	Sequence 205, App
C 18	15.6	86.7	713059	17	US-10-027-632-174581	Sequence 174581,
C 19	15.6	86.7	713059	13	US-10-027-632-174581	Sequence 174581,
C 20	15.6	86.7	2162598	21	US-10-472-928-4979	Sequence 4979, App
C 21	15.2	84.4	330	15	US-10-156-761-6136	Sequence 6136, App
C 22	14.6	81.1	110021	21	US-10-461-862-83	Sequence 83, App
C 23	14.6	81.1	209	9	US-09-294-0938-4374	Sequence 4374, App
C 24	14.6	81.1	271	20	US-10-425-115-76979	Sequence 76979, A
C 25	14.6	81.1	348	20	US-10-425-115-131598	Sequence 131598,
C 26	14.6	81.1	521	20	US-10-856-499-238	Sequence 238, App
C 27	14.6	81.1	521	20	US-10-856-499-1210	Sequence 1210, App
C 28	14.6	81.1	558	15	US-10-259-165-501	Sequence 501, App
C 29	14.6	81.1	735	19	US-10-437-963-73433	Sequence 73433, A
C 30	14.6	81.1	1041	17	US-10-107-431-2	Sequence 2, App1
C 31	14.6	81.1	1068	17	US-10-084-846A-53	Sequence 53, App1
C 32	14.6	81.1	1092	17	US-10-369-493-40179	Sequence 40179, A
C 33	14.6	81.1	1209	17	US-10-356-493-24261	Sequence 24261, A
C 34	14.6	81.1	1268	15	US-10-156-761-2078	Sequence 2078, App
C 35	14.6	81.1	1482	15	US-10-156-761-4032	Sequence 4032, App
C 36	14.6	81.1	1674	17	US-10-369-493-35833	Sequence 35833, A
C 37	14.6	81.1	1707	17	US-10-369-493-24890	Sequence 24890, A
C 38	14.6	81.1	1749	9	US-09-821-167-13	Sequence 13, App1
C 39	14.6	81.1	254	17	US-10-108-260A-1946	Sequence 1946, App
C 40	14.6	81.1	3013	19	US-10-437-963-40436	Sequence 40436, A
C 41	14.6	81.1	3799	17	US-10-369-493-26821	Sequence 26821, A
C 42	14.6	81.1	5874	17	US-10-365-743-119	Sequence 119, App
C 43	14.6	81.1	7564	19	US-10-476-397-15	Sequence 15, App
C 44	14.6	81.1	11188	9	US-09-821-167-1	Sequence 1, App1
C 45	14.6	81.1	14061	17	US-10-093-463-73	Sequence 73, App1
C 46	14.6	81.1	14109	17	US-10-093-463-71	Sequence 71, App1
C 47	14.6	81.1	17083	19	US-10-475-970-4	Sequence 4, App1
C 48	14.6	81.1	20499	20	US-10-723-860-5701	Sequence 5701, App
C 49	14.6	81.1	23907	13	US-10-077-130-6	Sequence 6, App1
C 50	14.6	81.1	2410	21	US-10-077-130-4	Sequence 4, App1
C 51	14.6	81.1	33825	13	US-10-866-089-1	Sequence 1, App1
C 52	14.6	81.1	33825	21	US-10-866-089-12	Sequence 12, App1
C 53	14.6	81.1	43865	19	US-10-322-289-119	Sequence 226, App
C 54	14.6	81.1	45055	17	US-10-107-431-27	Sequence 277, App
C 55	14.6	81.1	59816	17	US-10-084-846A-2	Sequence 2, App1
C 56	14.6	81.1	71222	13	US-10-084-846A-2	Sequence 1942, App
C 57	14.6	81.1	1236646	19	US-10-087-193-1942	Sequence 1, App1
C 58	14.6	81.1	2256646	19	US-10-470-565-1	Sequence 1, App1
C 59	14.6	81.1	2256646	15	US-10-156-761-1	Sequence 1, App1
C 60	14.6	80.0	144	16	US-10-012-952A-33	Sequence 33, App1
C 61	14.4	80.0	455	10	US-09-918-995-2309	Sequence 82244, A
C 62	14.4	80.0	4921	19	US-10-437-963-82224	Sequence 2309, App
C 63	14.4	80.0	1116	17	US-10-369-493-39328	Sequence 39328, A
C 64	14.4	80.0	1116	17	US-10-369-493-39703	Sequence 39703, A
C 65	14.4	80.0	1116	17	US-10-369-493-40056	Sequence 40056, A
C 66	14.4	80.0	1116	17	US-10-369-493-40056	Sequence 46606, A
C 67	14.4	80.0	1125	18	US-10-425-115-7140	Sequence 7140, App
C 68	14.4	80.0	1185	17	US-10-369-493-46606	Sequence 46606, A
C 69	14.4	80.0	1868	18	US-10-425-115-5724	Sequence 5724, App
C 70	14.4	80.0	2237	19	US-10-437-963-84314	Sequence 84314, A
C 71	14.4	80.0	4017	20	US-10-425-115-40887	Sequence 40887, A
C 72	14.4	80.0	4017	19	US-10-437-963-19901	Sequence 19901, A
C 73	14.4	80.0	4323	19	US-10-437-963-19897	Sequence 19897, A
C 74	14.4	80.0	3853122	21	US-10-741-600-17796	Sequence 17796, App
C 75	14.4	80.0	653122	23	US-10-087-193-226	Sequence 226, App
C 76	14.4	80.0	653122	20	US-10-087-193-226	Sequence 84144, A
C 77	14.2	78.9	253	10	US-10-425-115-84144	Sequence 226, App
C 78	14.2	78.9	471	9	US-09-934-903-9	Sequence 9, App1
C 79	14.2	78.9	471	9	US-09-934-868-65	Sequence 65, App1
C 80	14.2	78.9	471	10	US-09-941-947A-33	Sequence 13, App1

81	14.2	78.9	471	18	US-10-700-003-9	Sequence 9, Appli	C 154	14	77.8	564	20	US-10-425-115-138353	Sequence 138353,
82	14.2	78.9	471	18	US-10-363-567-13	Sequence 13, Appl	155	14	77.8	565	13	US-10-027-632-137998	Sequence 137998,
83	14.2	78.9	471	22	US-10-701-200-65	Sequence 65, Appl	156	14	77.8	565	13	US-10-027-632-137998	Sequence 137998,
84	14.2	78.9	474	15	US-10-156-761-4846	Sequence 4846, Ap	157	14	77.8	573	20	US-10-424-599-8518	Sequence 8518, Ap
85	14.2	78.9	575	20	US-10-425-115-37988	Sequence 37988, A	158	14	77.8	574	18	US-10-425-115-118455	Sequence 118455,
C 86	14.2	78.9	579	15	US-10-259-165-701	Sequence 701, App	159	14	77.8	588	15	US-10-156-761-3225	Sequence 3225, Ap
C 87	14.2	78.9	610	19	US-10-767-701-8566	Sequence 8566, Ap	160	14	77.8	604	9	US-10-767-701-659	Sequence 659, App
88	14.2	78.9	645	15	US-10-156-761-180	Sequence 180, App	C 161	14	77.8	618	9	US-09-770-149-608	Sequence 808, App
89	14.2	78.9	741	15	US-10-156-761-6450	Sequence 6450, App	C 162	14	77.8	637	13	US-10-027-632-220440	Sequence 220440,
90	14.2	78.9	796	15	US-10-062-831-44	Sequence 44, Appl	C 163	14	77.8	637	13	US-10-027-632-220441	Sequence 220441,
91	14.2	78.9	796	16	US-10-062-599-44	Sequence 44, Appl	C 164	14	77.8	637	13	US-10-027-632-220442	Sequence 220442,
92	14.2	78.9	897	17	US-10-369-483-39451	Sequence 39451, A	C 165	14	77.8	637	13	US-10-027-632-220443	Sequence 220443,
93	14.2	78.9	897	17	US-10-369-483-39835	Sequence 39835, A	C 166	14	77.8	637	17	US-10-027-632-220440	Sequence 220440,
94	14.2	78.9	909	17	US-10-369-493-39835	Sequence 39835, A	C 167	14	77.8	637	17	US-10-027-632-220441	Sequence 220441,
95	14.2	78.9	909	17	US-10-369-493-39835	Sequence 39835, A	C 168	14	77.8	637	17	US-10-027-632-220442	Sequence 220442,
96	14.2	78.9	1179	21	US-10-200-545-18	Sequence 18, Appl	C 169	14	77.8	637	13	US-10-027-632-131009	Sequence 131009,
C 97	14.2	78.9	1295	14	US-10-281-024-2	Sequence 2, Appli	170	14	77.8	660	17	US-10-027-632-131009	Sequence 131009,
98	14.2	78.9	1504	20	US-10-425-115-139009	Sequence 139009,	171	14	77.8	660	17	US-10-027-632-131009	Sequence 131009,
99	14.2	78.9	1836	17	US-10-369-483-39783	Sequence 39783, A	C 172	14	77.8	661	19	US-10-767-701-3355	Sequence 3355, Ap
100	14.2	78.9	2035	17	US-10-369-483-39783	Sequence 39783, A	C 173	14	77.8	666	9	US-09-770-149-445	Sequence 445, App
101	14.2	78.9	2035	17	US-10-369-493-39026	Sequence 39026, A	C 174	14	77.8	669	19	US-10-437-963-54737	Sequence 54737, A
102	14.2	78.9	2078	19	US-10-437-963-71557	Sequence 39393, A	C 175	14	77.8	671	10	US-09-952-060-13	Sequence 13, Appl
C 103	14.2	78.9	2343	20	US-10-437-963-71557	Sequence 71557, A	C 176	14	77.8	671	14	US-10-149-640-5	Sequence 5, Appl
104	14.2	78.9	2343	20	US-10-437-963-71557	Sequence 10168, A	C 177	14	77.8	671	14	US-10-380-641-13	Sequence 13, Appl
C 105	14.2	78.9	2627	17	US-10-369-483-37750	Sequence 37750, A	C 178	14	77.8	671	21	US-10-636-730-13	Sequence 13, Appl
106	14.2	78.9	4122	19	US-10-437-963-56914	Sequence 56914, A	C 179	14	77.8	671	15	US-10-156-761-7519	Sequence 7519, Ap
107	14.2	78.9	25	21	US-10-719-900-452264	Sequence 452264,	180	14	77.8	684	15	US-10-425-114-31951	Sequence 31951, A
C 108	14.2	78.9	25	21	US-10-719-900-452265	Sequence 452265,	181	14	77.8	705	20	US-10-425-115-75219	Sequence 75219, A
C 109	14.2	78.9	60	10	US-09-804-980-200	Sequence 200, App	182	14	77.8	728	20	US-10-425-115-11625	Sequence 11625,
C 110	14.2	78.9	139	14	US-10-149-640-28	Sequence 202, App	C 183	14	77.8	740	19	US-10-437-963-93255	Sequence 93255, A
C 111	14.2	78.9	255	17	US-10-084-846A-13	Sequence 20, Appl	184	14	77.8	763	17	US-10-139-794-210	Sequence 210, App
C 112	14.2	78.9	256	17	US-10-425-115-110668	Sequence 13, Appl	185	14	77.8	768	20	US-10-425-115-9753	Sequence 9753, App
113	14.2	78.9	268	9	US-09-922-217-723	Sequence 110668,	186	14	77.8	789	17	US-10-369-493-35141	Sequence 35141, A
114	14.2	78.9	268	9	US-09-833-263-723	Sequence 723, App	187	14	77.8	789	17	US-10-369-493-38322	Sequence 38322, A
115	14.2	78.9	268	13	US-10-025-380-773	Sequence 723, App	188	14	77.8	789	17	US-10-369-493-38784	Sequence 38784, A
C 116	14.2	78.9	291	10	US-09-804-980-174	Sequence 174, App	189	14	77.8	801	20	US-10-425-115-153290	Sequence 153290,
C 117	14.2	78.9	291	10	US-09-804-980-198	Sequence 198, App	C 190	14	77.8	831	20	US-10-425-115-85980	Sequence 85980, A
C 118	14.2	78.9	299	20	US-10-425-115-72578	Sequence 72578, A	191	14	77.8	836	20	US-10-425-115-85984	Sequence 85984, A
C 119	14.2	78.9	303	17	US-10-282-122A-19843	Sequence 19843, A	192	14	77.8	840	17	US-10-369-493-35515	Sequence 35515, A
120	14.2	78.9	311	19	US-10-767-701-17402	Sequence 17402, A	193	14	77.8	868	20	US-10-425-115-98215	Sequence 98215, A
121	14.2	78.9	312	17	US-10-243-535A-47073	Sequence 47073, A	194	14	77.8	876	19	US-10-767-701-10747	Sequence 10747, A
122	14.2	78.9	312	18	US-10-085-783A-47073	Sequence 47073, A	195	14	77.8	895	18	US-10-424-599-37600	Sequence 37600, A
123	14.2	78.9	322	18	US-10-424-599-102046	Sequence 102046,	196	14	77.8	913	18	US-10-425-114-22860	Sequence 22860, A
C 124	14.2	78.9	323	11	US-09-978-360A-188	Sequence 188, App	197	14	77.8	918	18	US-10-425-114-28280	Sequence 28280, A
125	14.2	78.9	336	19	US-10-437-963-84888	Sequence 84888, A	198	14	77.8	923	18	US-10-425-114-28312	Sequence 28312, A
C 126	14.2	78.9	336	20	US-10-425-115-164181	Sequence 164181,	199	14	77.8	933	18	US-10-425-114-28318	Sequence 28318, A
C 127	14.2	78.9	343	20	US-10-425-115-36422	Sequence 36422, A	200	14	77.8	945	18	US-10-424-599-91985	Sequence 91985, A
C 128	14.2	78.9	348	10	US-09-803-719-92	Sequence 92, Appl	201	14	77.8	945	18	US-10-425-114-28339	Sequence 28339, A
C 129	14.2	78.9	354	19	US-10-474-495-64	Sequence 64, Appl	202	14	77.8	948	20	US-10-425-115-70627	Sequence 70627, A
C 130	14.2	78.9	356	11	US-09-864-408A-3221	Sequence 3221, App	203	14	77.8	949	18	US-10-425-114-28554	Sequence 28554, A
131	14.2	78.9	364	20	US-10-425-115-89605	Sequence 89605, A	204	14	77.8	951	18	US-10-425-114-28178	Sequence 28178, A
C 132	14.2	78.9	390	20	US-10-425-115-20563	Sequence 20563, A	205	14	77.8	951	18	US-10-425-114-28178	Sequence 28178, A
C 133	14.2	78.9	413	10	US-09-918-995-4176	Sequence 4176, Ap	206	14	77.8	953	18	US-10-425-114-28178	Sequence 28178, A
C 134	14.2	78.9	416	18	US-10-424-599-36816	Sequence 36816, Ap	207	14	77.8	958	18	US-10-425-114-28193	Sequence 28193, A
C 135	14.2	78.9	428	10	US-09-918-995-16713	Sequence 16713, A	208	14	77.8	971	18	US-10-425-114-15752	Sequence 15752, A
C 136	14.2	78.9	428	19	US-10-437-963-14432	Sequence 14432, A	C 209	14	77.8	979	18	US-10-240-140A-117	Sequence 117, App
137	14.2	78.9	449	20	US-10-425-115-67593	Sequence 67593, A	210	14	77.8	1041	17	US-10-107-431-12	Sequence 12, Appl
138	14.2	78.9	451	9	US-09-960-352-1342	Sequence 1342, App	211	14	77.8	1041	17	US-10-107-431-12	Sequence 12, Appl
C 139	14.2	78.9	458	20	US-10-425-115-38190	Sequence 38190, A	212	14	77.8	1047	17	US-10-084-846A-101	Sequence 846A-101
C 140	14.2	78.9	467	10	US-09-918-995-30974	Sequence 30974, A	213	14	77.8	1047	17	US-10-084-846A-101	Sequence 846A-101
C 141	14.2	78.9	469	10	US-09-918-995-32468	Sequence 32468, A	214	14	77.8	1049	21	US-10-487-901-2188	Sequence 2188, App
C 142	14.2	78.9	469	20	US-10-425-115-18769	Sequence 18769, A	215	14	77.8	1059	20	US-10-425-115-85985	Sequence 85985, A
C 143	14.2	78.9	477	17	US-10-282-122A-6982	Sequence 6982, Ap	216	14	77.8	1062	11	US-09-758-759-32	Sequence 32, Appl
144	14.2	78.9	479	20	US-10-425-115-20558	Sequence 20558, A	217	14	77.8	1063	19	US-10-767-701-14717	Sequence 14717, A
145	14.2	78.9	482	13	US-10-027-632-278807	Sequence 278807,	218	14	77.8	1078	18	US-10-425-114-5266	Sequence 5266, Ap
146	14.2	78.9	482	17	US-10-027-632-278807	Sequence 278807,	C 219	14	77.8	1089	17	US-10-084-846A-9	Sequence 9, Appli
C 147	14.2	78.9	492	19	US-10-767-701-5969	Sequence 5969, Ap	220	14	77.8	1124	21	US-10-487-901-6180	Sequence 6180, App
C 148	14.2	78.9	519	20	US-10-425-115-86591	Sequence 86591, A	221	14	77.8	1146	17	US-10-369-493-47075	Sequence 47075, A
C 149	14.2	78.9	538	20	US-10-425-115-28191	Sequence 28191, A	222	14	77.8	1156	20	US-10-425-115-85991	Sequence 85991, A
C 150	14.2	78.9	539	20	US-10-425-115-152949	Sequence 152949,	C 223	14	77.8	1167	10	US-09-933-167-107	Sequence 107, App
151	14.2	78.9	540	9	US-09-854-133-658	Sequence 658, App	224	14	77.8	1167	14	US-10-004-860-107	Sequence 107, App
152	14.2	78.9	540	15	US-10-144-649A-658	Sequence 658, App	225	14	77.8	1167	14	US-10-023-282-107	Sequence 107, App
153	14.2	78.9	550	20	US-10-425-115-20560	Sequence 20560, A	C 226	14	77.8	1179	16	US-10-091-007-223	Sequence 223, App

C 227	14	77.8	1182	17	US-10-369-493-2447	Sequence 2427, A
C 228	14	77.8	1200	20	US-10-425-115-20566	Sequence 20566, A
C 229	14	77.8	1218	16	US-10-174-209-25	Sequence 25, Appl
C 230	14	77.8	1227	15	US-10-156-761-3900	Sequence 3900, Ap
C 231	14	77.8	1235	19	US-10-437-963-86392	Sequence 86392, A
C 232	14	77.8	1266	17	US-10-369-493-45504	Sequence 45504, A
C 233	14	77.8	1295	19	US-10-437-963-9754	Sequence 9754, Ap
C 234	14	77.8	1308	17	US-10-282-122A-33059	Sequence 33059, A
C 235	14	77.8	1319	20	US-10-437-963-68782	Sequence 68782, A
C 236	14	77.8	1326	19	US-10-425-115-175783	Sequence 175783, A
C 237	14	77.8	1335	18	US-10-425-114-14047	Sequence 14047, A
C 238	14	77.8	1341	18	US-10-425-114-32758	Sequence 32758, A
C 239	14	77.8	1356	9	US-09-938-330-1	Sequence 1, Appl
C 240	14	77.8	1359	19	US-10-437-963-20972	Sequence 20972, A
C 241	14	77.8	1362	18	US-10-425-114-35232	Sequence 35232, A
C 242	14	77.8	1377	20	US-10-437-963-64659	Sequence 64659, A
C 243	14	77.8	1377	20	US-10-411-910A-266	Sequence 266, App
C 244	14	77.8	1381	20	US-10-425-115-129730	Sequence 129730, A
C 245	14	77.8	1386	17	US-10-369-493-31705	Sequence 31705, A
C 246	14	77.8	1389	17	US-10-417-700A-110	Sequence 110, App
C 247	14	77.8	1392	18	US-10-417-700A-104	Sequence 104, App
C 248	14	77.8	1395	18	US-10-417-700A-108	Sequence 108, App
C 249	14	77.8	1410	18	US-10-417-700A-106	Sequence 106, App
C 250	14	77.8	1412	20	US-10-425-115-16343	Sequence 16343, A
C 251	14	77.8	1440	9	US-09-712-363-94	Sequence 94, Appl
C 252	14	77.8	1440	17	US-10-282-122A-28525	Sequence 28525, A
C 253	14	77.8	1452	17	US-10-282-122A-26564	Sequence 26564, A
C 254	14	77.8	1461	9	US-09-938-330-5	Sequence 5, Appl
C 255	14	77.8	1476	17	US-10-282-122A-28799	Sequence 28799, A
C 256	14	77.8	1479	15	US-10-156-761-1451	Sequence 1451, Ap
C 257	14	77.8	1482	17	US-10-282-122A-11680	Sequence 11680, A
C 258	14	77.8	1488	21	US-10-490-913-3	Sequence 3, Appl
C 259	14	77.8	1488	21	US-10-887-553A-756	Sequence 756, App
C 260	14	77.8	1488	21	US-10-956-157-1639	Sequence 1639, App
C 261	14	77.8	1488	21	US-10-956-157-1639	Sequence 1639, A
C 262	14	77.8	1491	19	US-10-437-963-55898	Sequence 55898, A
C 263	14	77.8	1494	17	US-10-282-122A-28783	Sequence 28783, A
C 264	14	77.8	1503	19	US-10-437-963-13836	Sequence 13836, A
C 265	14	77.8	1503	21	US-10-965-898-120	Sequence 120, App
C 266	14	77.8	1524	16	US-10-304-220-5	Sequence 5, Appl
C 267	14	77.8	1530	19	US-10-437-963-47467	Sequence 47467, A
C 268	14	77.8	1559	17	US-10-282-122A-19326	Sequence 19326, A
C 269	14	77.8	1563	21	US-10-956-157-4211	Sequence 4211, Ap
C 270	14	77.8	1593	17	US-10-282-122A-25611	Sequence 25611, A
C 271	14	77.8	1597	19	US-10-437-963-20809	Sequence 20809, A
C 272	14	77.8	1616	20	US-10-425-115-180527	Sequence 180527, A
C 273	14	77.8	1660	17	US-10-415-378-27	Sequence 27, Appl
C 274	14	77.8	1680	9	US-09-974-300-2326	Sequence 2326, Ap
C 275	14	77.8	1686	18	US-10-425-114-32782	Sequence 32782, A
C 276	14	77.8	1704	15	US-10-156-761-4987	Sequence 4987, A
C 277	14	77.8	1734	15	US-10-156-761-29	Sequence 29, Appl
C 278	14	77.8	1740	20	US-10-425-115-85990	Sequence 85990, A
C 279	14	77.8	1811	9	US-09-822-830A-617	Sequence 617, App
C 280	14	77.8	1832	18	US-10-425-114-28132	Sequence 28132, A
C 281	14	77.8	1838	19	US-10-437-963-88815	Sequence 88815, A
C 282	14	77.8	1857	17	US-10-282-122A-15065	Sequence 15065, A
C 283	14	77.8	1866	14	US-10-166-087-13	Sequence 13, Appl
C 284	14	77.8	1879	18	US-10-425-114-32261	Sequence 32261, A
C 285	14	77.8	1879	18	US-10-641-643-742	Sequence 742, App
C 286	14	77.8	1882	17	US-10-108-260A-1035	Sequence 1035, Ap
C 287	14	77.8	1963	20	US-10-739-930-4011	Sequence 4011, Ap
C 288	14	77.8	2055	18	US-10-425-114-29541	Sequence 29541, A
C 289	14	77.8	2058	20	US-10-425-115-85988	Sequence 85988, A
C 290	14	77.8	2103	15	US-10-156-761-4491	Sequence 4491, Ap
C 291	14	77.8	2103	17	US-10-282-122A-25741	Sequence 25741, A
C 292	14	77.8	2106	20	US-10-425-115-46880	Sequence 46880, A
C 293	14	77.8	2166	17	US-10-330-011-127	Sequence 127, App
C 294	14	77.8	2176	17	US-10-264-237-994	Sequence 994, App
C 295	14	77.8	2187	18	US-10-425-114-30640	Sequence 30640, A
C 296	14	77.8	2212	14	US-10-424-599-72496	Sequence 72496, A
C 297	14	77.8	2226	20	US-10-425-115-47014	Sequence 47014, A
C 298	14	77.8	2334	17	US-10-260-238-1133	Sequence 1133, Ap
C 299	14	77.8	2367	19	US-10-437-963-43182	Sequence 43182, A

300 14 77.8 2394 17 US-10-369-493-26878 Sequence 26878, A

## ALIGNMENTS

## RESULT 1

US-10-017-471A-1  
 ; Sequence 1, Application US/10017471A  
 ; Publication No. US20030124644A1

GENERAL INFORMATION:  
 ; APPLICANT: Takano, Eiko

APPLICANT: Bibb, Mervyn  
 ; TITLE OF INVENTION: Antibiotic Production

FILE REFERENCE: 0380-P02329U51  
 ; CURRENT APPLICATION NUMBER: US/10/017,471A

PRIOR FILING DATE: 2001-10-23  
 ; PRIOR APPLICATION NUMBER: US 60/242,561

PRIOR FILING DATE: 2000-10-23  
 ; SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 1  
 ; LENGTH: 18

TYPE: DNA  
 ; ORGANISM: Artificial Sequence

FEATURES:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Synthetic

OTHER INFORMATION: oligonucleotide  
 ; US-10-017-471A-1

Query Match 95.6%; Score 17.2; DB 15; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 95.6%; Score 17.2; DB 15; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 95.6%; Score 17.2; DB 15; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 95.6%; Score 17.2; DB 15; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 95.6%; Score 17.2; DB 15; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 95.6%; Score 17.2; DB 15; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 95.6%; Score 17.2; DB 15; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 95.6%; Score 17.2; DB 15; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 95.6%; Score 17.2; DB 15; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 95.6%; Score 17.2; DB 15; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 95.6%; Score 17.2; DB 15; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 95.6%; Score 17.2; DB 15; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1 GACCACGTSCCGGCATG 18
         |||||:|:|||||
Db      160 GACCACGTCCCCGGCATG 177

```

```

RESULT 3
US-10-156-761-1/c
? Sequence 1, Application US/10156761
? Publication No. US20030119018A1
? GENERAL INFORMATION:
? APPLICANT: OMIEDA, SATOSHI
? APPLICANT: IKEDA, HARUO
? APPLICANT: ISHIKAWA, JIN
? APPLICANT: HOKIKAWA, HIROSHI
? APPLICANT: SHIBA, TADAYOSHI
? APPLICANT: SAKAKI, YOSHIYUKI
? APPLICANT: HATTORI, MASHIARA
? TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
? FILE REFERENCE: 249-262
? CURRENT APPLICATION NUMBER: US/10/156,761
? CURRENT FILING DATE: 2002-05-29
? PRIOR APPLICATION NUMBER: JP 2001-204089
? PRIOR FILING DATE: 2001-05-30
? PRIOR APPLICATION NUMBER: JP 2001-272697
? PRIOR FILING DATE: 2001-08-02
? NUMBER OF SEQ ID NOS: 15109
? SEQ ID NO 1
? LENGTH: 9025608
? TYPE: DNA
? ORGANISM: Streptomyces avermitilis
? FEATURE:
? NAME/KEY: misc feature
? LOCATION: (4187715)
? OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

```

	Query Match	95.6%	Score 17.2;	DB 15;	Length 9025608;
	Best Local Similarity	88.9%	Pred. No. 6.8;		
	Matches	16;	Conservative	2;	Mismatches 0;
					Indels 0;
	Gaps	0;			
QY	1 GACCACGTSCCGGCATG	18			
	:				
Db	4361014 GACCACTGTCCCGGCATG	4360997			

```

1  RESULT 4
2  US-10-017-471A-19/c
3  : Sequence 19, Application US/10017471A
4  : Publication No. US20030124644A1
5  : GENERAL INFORMATION:
6  : APPLICANT: Takano, Eriko
7  : APPLICANT: Bibb, Meryn
8  : TITLE OF INVENTION: Antibiotic Production
9  : FILE REFERENCE: US80-P02329US1
10 : CURRENT APPLICATION NUMBER: US/10/017,471A
11 : CURRENT FILING DATE: 2001-10-23
12 : PRIOR APPLICATION NUMBER: 2001-10-23
13 : PRIOR FILING DATE: 2000-10-23
14 : NUMBER OF SEQ ID NOS: 19
15 : SOFTWARE: PatentIn Ver. 2.1
16 : SEQ ID NO 19
17 : LENGTH: 4346
18 : TYPE: DNA
19 : ORGANISM: Streptomyces coelicolor
20 : US-10-017-471A-19

```

Query Match	88.9%	Score 16;	DB 15;	Length 4346;
Best Local Similarity	88.9%	Pred. No. 1.2e+02;		
Matches 16;	Conservative 1;	Mismatches 0;	Gaps 0;	

```

RESULT 5
US-09-975-719-155
; Sequence 155, Application US/09975719
; Publication No. US20030022349A1
; GENERAL INFORMATION:
; APPLICANT: Anshel, Frederick M.
; APPLICANT: Rahme, Lawrence G.
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361003
; CURRENT APPLICATION NUMBER: US/09/975, 719
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 09/199,637
; PRIOR FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: US 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 155
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-975-719-155

```

Query Match	86.7%	Score 15.6	DB 10	Length 513
Best Local Similarity	83.3%	Pred. No. 2.9e+02		
Matches 15, Conservative	2	Mismatches 1	Indels 0	Gaps 0

RESULT 6  
 US-10-617-320-417/c  
 ; Sequence 417, Application US/10617320  
 ; Publication No. US20050136404A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn A Doucette-Stamm and David Bush  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
 ; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
 ; THERAPEUTICS  
 ;  
 ; NUMBER OF SEQUENCES: 5206  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
 ; STREET: 100 Beaver Street  
 ; CITY: Waltham  
 ; STATE: Massachusetts  
 ; COUNTRY: USA  
 ; ZIP: 02354  
 ;  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: CD/ROM ISO9660  
 ; COMPUTER: <Unknown>  
 ; OPERATING SYSTEM: <Unknown>  
 ; SOFTWARE: <Unknown>  
 ;  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/617,320  
 ; FILING DATE: 10-Jul-2003  
 ;  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/107,433  
 ; FILING DATE: 30-Jun-1998  
 ; APPLICATION NUMBER: 60/ 085131  
 ; FILING DATE: May 12, 1998  
 ; APPLICATION NUMBER: 60/051553  
 ; FILING DATE: July 2, 1997  
 ;  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Ariniello, Pamela Deneke  
 ; REGISTRATION NUMBER: 40,489  
 ; REFERENCE/DOCKET NUMBER: GTC--011  
 ;  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (781) 893-5007

```

; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 417:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 612 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
;   ORGANISM: Streptococcus pneumoniae
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: (b) LOCATION 1...612
;     SEQUENCE DESCRIPTION: SEQ ID NO: 417:
US-10-617-320-417

Query Match      86.7%; Score 15.6; DB 22; Length 612;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 GACCACTGTCGCGCATG 18
        |||||:|:|:|
Db      160 GACCACTGTCGCGCATG 143

RESULT 7
US-09-975-719-153/c
; Sequence 153, Application US/09975719
; Publication No. US2003002349A1
; GENERAL INFORMATION:
;   APPLICANT: Ausubel, Frederick M.
;   APPLICANT: Rahme, Laurence G.
;   TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
;   FILE REFERENCE: 00786/351003
;   CURRENT APPLICATION NUMBER: US/09/975,719
;   PRIOR FILING DATE: 2001-10-10
;   PRIOR APPLICATION NUMBER: US 09/199,637
;   PRIOR FILING DATE: 1998-11-25
;   PRIOR APPLICATION NUMBER: US 60/066,517
;   PRIOR FILING DATE: 1997-11-25
;   NUMBER OF SEQ ID NOS: 437
;   SOFTWARE: FastSeq for Windows Version 4.0
;   SEQ ID NO 153
;   LENGTH: 762
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-975-719-153

Query Match      86.7%; Score 15.6; DB 10; Length 762;
Best Local Similarity 83.3%; Pred. No. 2.7e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 GACCACTGTCGCGCATG 18
        |||||:|:|:|
Db      303 GACCACTGTCGCGCATG 286

RESULT 8
US-10-027-632-100941/c
; Sequence 100941, Application US/10027632
; Publication No. US2002019837A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;   FILE REFERENCE: 108827.129
;   CURRENT APPLICATION NUMBER: US/10/027,632
;   CURRENT FILING DATE: 2002-04-30
;   PRIOR APPLICATION NUMBER: US 60/146,002
;   PRIOR FILING DATE: 1999-08-09
;   NUMBER OF SEQ ID NOS: 325720
;   SOFTWARE: FastSeq for Windows Version 4.0
;   SEQ ID NO 100941
;   LENGTH: 1080
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-100941

Query Match      86.7%; Score 15.6; DB 13; Length 1080;
Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 GACCACTGTCGCGCATG 18
        |||||:|:|:|
Db      206 GACCACTGTCGCGCATG 189

RESULT 10
US-10-027-632-100943/c
```

```

; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 100941
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-100941

Query Match      86.7%; Score 15.6; DB 13; Length 1080;
Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 GACCACTGTCGCGCATG 18
        |||||:|:~|~|
Db      206 GACCACTGTCGCGCATG 189

RESULT 9
US-10-027-632-100942/c
; Sequence 100942, Application US/10027632
; Publication No. US2002019837A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;   FILE REFERENCE: 108827.129
;   CURRENT APPLICATION NUMBER: US/10/027,632
;   CURRENT FILING DATE: 2002-04-30
;   PRIOR APPLICATION NUMBER: US 60/218,006
;   PRIOR FILING DATE: 2000-07-12
;   PRIOR APPLICATION NUMBER: US 60/198,676
;   PRIOR FILING DATE: 2000-04-20
;   PRIOR APPLICATION NUMBER: US 60/193,483
;   PRIOR FILING DATE: 2000-03-29
;   PRIOR APPLICATION NUMBER: US 60/185,218
;   PRIOR FILING DATE: 2000-02-24
;   PRIOR APPLICATION NUMBER: US 60/167,363
;   PRIOR FILING DATE: 1999-11-23
;   PRIOR APPLICATION NUMBER: US 60/156,358
;   PRIOR FILING DATE: 1999-09-28
;   PRIOR APPLICATION NUMBER: US 60/146,002
;   PRIOR FILING DATE: 1999-08-09
;   NUMBER OF SEQ ID NOS: 325720
;   SOFTWARE: FastSeq for Windows Version 4.0
;   SEQ ID NO 100942
;   LENGTH: 1080
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-100942
```



```
; Sequence 100943, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 100943
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-100943

Query Match      86.7%; Score 15.6; DB 13; Length 1080;
Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GACCACGTCCCGGCATG 18
      |||||:|:|:|:|:|
Db      206 GACCACGTCCCGGCAGG 189

RESULT 11
US-10-027-632-100941/c
; Sequence 100941, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 100941
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-100941
```

```
Query Match      86.7%; Score 15.6; DB 17; Length 1080;
Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GACCACGTCCCGGCATG 18
      |||||:|:|:|:|:|
Db      206 GACCACGTCCCGGCAGG 189

RESULT 12
US-10-027-632-100942/c
; Sequence 100942, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 100942
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-100942

Query Match      86.7%; Score 15.6; DB 17; Length 1080;
Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 GACCACGTCCCGGCATG 18
      |||||:|:~|:|:|:|
Db      206 GACCACGTCCCGGCAGG 189

RESULT 13
US-10-027-632-100943/c
; Sequence 100943, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
```

PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 100943  
LENGTH: 1080  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-100943

Query Match 86.7%; Score 15.6; DB 17; Length 1080;  
Best Local Similarity 83.3%; Pred. No. 2.5e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGSCCGGCGATG 18  
|||||:|||||  
Db 206 GACCACTGSCCGGCGATG 189

RESULT 14  
US-10-472-928-473/c  
Sequence 473, Application US/10472928  
Publication No. US2005020813A1  
GENERAL INFORMATION:  
APPLICANT: CHIRON SPA  
TITLE OF INVENTION: THE INSTITUTE FOR GENOMIC RESEARCH  
FILE REFERENCE: P026926MO  
CURRENT APPLICATION NUMBER: US/10/472,928  
CURRENT FILING DATE: 2003-09-26  
PRIOR APPLICATION NUMBER: GB-0107658.7  
PRIOR FILING DATE: 2001-03-27  
NUMBER OF SEQ ID NOS: 4979  
SOFTWARE: SeqWin99, version 1.03  
SEQ ID NO 473  
LENGTH: 1188  
TYPE: DNA  
ORGANISM: Streptococcus pneumoniae  
US-10-472-928-473

Query Match 86.7%; Score 15.6; DB 21; Length 1188;  
Best Local Similarity 83.3%; Pred. No. 2.4e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGSCCGGCGATG 18  
|||||:|||||  
Db 739 GACCACTGSCCGGCGATG 722

RESULT 15  
US-09-975-719-136  
Sequence 136, Application US/09975719  
Publication No. US20030022349A1  
GENERAL INFORMATION:  
APPLICANT: Ausubel, Frederick M.  
TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
FILE REFERENCE: 00786/361003  
CURRENT APPLICATION NUMBER: US/09/975,719  
CURRENT FILING DATE: 2001-10-10  
PRIOR APPLICATION NUMBER: US 09/199,637  
PRIOR FILING DATE: 1998-11-25  
PRIOR APPLICATION NUMBER: US 60/066,517  
PRIOR FILING DATE: 1997-11-25  
NUMBER OF SEQ ID NOS: 437  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 136  
LENGTH: 2048  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-975-719-136

Query Match 86.7%; Score 15.6; DB 10; Length 2048;  
Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGSCCGGCGATG 18  
|||||:|||||  
Db 1069 GACCACTGSCCGGCGATG 1086

RESULT 16  
US-08-961-527-205/c  
Sequence 205, Application US/08961527  
Publication No. US20020032323A1  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/961,527  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 205:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2395 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-205

Query Match 86.7%; Score 15.6; DB 8; Length 2395;  
Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGSCCGGCGATG 18  
|||||:|||||  
Db 1257 GACCACTGSCCGGCGATG 1240

RESULT 17  
US-10-158-844-205/c  
Sequence 205, Application US/10158844  
Publication No. US20040029118A1  
GENERAL INFORMATION:  
APPLICANT: Kunsch et al.  
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
NUMBER OF SEQUENCES: 391  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville

```
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-R
COMPUTER: Dell Latitude Pentium 3
OPERATING SYSTEM: Windows 98
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/158.844
FILING DATE: 03-Jun-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/961,527
FILING DATE: 1997-10-30
APPLICATION NUMBER: US 60/029,960
FILING DATE: 1996-10-31
ATTORNEY/AGENT INFORMATION:
NAME: Hyman, Mark J.
REGISTRATION NUMBER: 46,789
REFERENCE/DOCKET NUMBER: PB340P1D1
INFORMATION FOR SEQ ID NO: 205:
SEQUENCE CHARACTERISTICS:
LENGTH: 2395 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 205:
US-10-158-844-205

Query Match      86.7%; Score 15.6; DB 17; Length 2395;
Best Local Similarity 83.3%; Pred. No. 2,1e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Cy 1 GACCAGTSCCGGCATG 18
Db 1257 GACCAGTSCCGGCATG 1240

RESULT 18
US-10-027-632-174581
Sequence 174581, Application US/10027632
Publication No. US20020198371A1
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 174581
LENGTH: 713059
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(713059)
OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174581
```

```
OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174581

Query Match      86.7%; Score 15.6; DB 13; Length 713059;
Best Local Similarity 83.3%; Pred. No. 69;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Cy 1 GACCAGTSCCGGCATG 18
Db 462752 GACCAGTSCCGGCATG 462769

RESULT 19
US-10-027-632-174581
Sequence 174581, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 174581
LENGTH: 713059
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(713059)
OTHER INFORMATION: n = A,T,C or G
US-10-027-632-174581

Query Match      86.7%; Score 15.6; DB 17; Length 713059;
Best Local Similarity 83.3%; Pred. No. 69;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Cy 1 GACCAGTSCCGGCATG 18
Db 462752 GACCAGTSCCGGCATG 462769

RESULT 20
US-10-472-928-4979/c
Sequence 4979, Application US/10472928
Publication No. US20050020813A1
GENERAL INFORMATION:
APPLICANT: CHIRON SPA
TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
FILE REFERENCE: P026926W0
CURRENT APPLICATION NUMBER: US/10/472,928
CURRENT FILING DATE: 2003-09-26
PRIOR APPLICATION NUMBER: GB-0107658.7
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 4979
SOFTWARE: SeqWin99, version 1.03
```

SEQ ID NO 4979  
LENGTH: 2162598  
TYPE: DNA  
ORGANISM: Streptococcus pneumoniae  
US-10-472-928-4979

Query Match 86.7%; Score 15.6; DB 21; Length 2162598;  
Best Local Similarity 83.3%; Pred. No. 55;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGSCGCGCATG 18  
|||||:|||||  
Db 190126 GACCACTGCGCGCATG 190109

RESULT 21  
US-10-156-761-6136/c  
Sequence 6136, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 6136  
LENGTH: 330  
TYPE: DNA  
ORGANISM: Streptococcus avermitilis  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(330)  
US-10-156-761-6136

Query Match 84.4%; Score 15.2; DB 15; Length 330;  
Best Local Similarity 87.5%; Pred. No. 4.9e+02;  
Matches 14; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCACGTSCGCGCATG 18  
|||||:|||||  
Db 193 CCACGTCCCGGCATG 178

RESULT 22  
US-10-461-862-83/c  
Sequence 83, Application US/10461862  
Publication No. US2005090343A1  
GENERAL INFORMATION:  
APPLICANT: David W. Morris  
APPLICANT: Marc S. Malandro  
TITLE OF INVENTION: Novel Therapeutic Targets in Cancer  
FILE REFERENCE: 529452001800  
CURRENT APPLICATION NUMBER: US/10/461,862  
CURRENT FILING DATE: 2003-06-13  
NUMBER OF SEQ ID NOS: 184  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 83  
LENGTH: 110021  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: misc\_feature

LOCATION: (1)...(110021)  
OTHER INFORMATION: n = A,T,C or G  
US-10-461-862-83

Query Match 84.4%; Score 15.2; DB 21; Length 110021;  
Best Local Similarity 87.5%; Pred. No. 1.6e+02;  
Matches 14; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 CCACGTSCGCGCATG 18  
|||||:|||||  
Db 41825 CCACGTCCCGGCATG 41810

RESULT 23  
US-09-294-093B-4374  
Sequence 4374, Application US/09294093B  
Patent No. US2001005135A1  
GENERAL INFORMATION:  
APPLICANT: Ito, Laura, Y.  
APPLICANT: Sherman, Bradley, K.  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL  
FILE REFERENCE: PL-0009 US  
CURRENT APPLICATION NUMBER: US/09/294,093B  
CURRENT FILING DATE: 1999-04-16  
PRIOR APPLICATION NUMBER: 60/082,567  
PRIOR FILING DATE: April 21, 1998  
NUMBER OF SEQ ID NOS: 6207  
SOFTWARE: PERL Program  
SEQ ID NO 4374  
LENGTH: 209  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: misc\_feature  
OTHER INFORMATION: Incyte ID No. US2001005135A1 700353928H1  
LOCATION: 36, 63, 114, 134  
OTHER INFORMATION: a, t, c, g, or other  
US-09-294-093B-4374

Query Match 81.1%; Score 14.6; DB 9; Length 209;  
Best Local Similarity 82.4%; Pred. No. 1.1e+03;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGSCGCGCAT 17  
|||||:|||||  
Db 87 GACTACGTCCCGGCAT 103

RESULT 24  
US-10-425-115-76979  
Sequence 76979, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 76979  
LENGTH: 271  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: M714577\_170231C.1  
US-10-425-115-76979



LOCATION: (478)..(478)  
OTHER INFORMATION: n = any nucleotide  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (528)..(528)  
OTHER INFORMATION: n = any nucleotide  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (539)..(539)  
OTHER INFORMATION: n = any nucleotide  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (553)..(553)  
OTHER INFORMATION: n = any nucleotide  
US-10-259-165-501

Query Match 81.1%; Score 14.6; DB 15; Length 568;  
Best Local Similarity 82.4%; Pred. NO. 8.8e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTSCCGGCAT 17  
|||||:|||||  
DB 378 GACCACTCCCGGCT 394

RESULT 29  
US-10-437-963-73433  
Sequence 73433, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovacic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21 (53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 73433  
LENGTH: 735  
TYPE: DNA  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_73714C.1  
US-10-437-963-73433

Query Match 81.1%; Score 14.6; DB 19; Length 735;  
Best Local Similarity 82.4%; Pred. NO. 8.4e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACCACGTSCCGGCATG 18  
|||||:|||||  
DB 317 ACCACGTCCCGGCTG 333

RESULT 30  
US-10-107-431-2  
Sequence 2, Application US/10107431  
Publication No. US20030224364A1  
GENERAL INFORMATION:  
APPLICANT: Farnet, Chris  
APPLICANT: Seafie, Alfredo  
APPLICANT: Zazopoulos, Emmanuel  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO  
TITLE OF INVENTION: BIOSYNTHETIC LOCI  
FILE REFERENCE: 3001-705  
CURRENT APPLICATION NUMBER: US/10/107,431

CURRENT FILING DATE: 2002-03-28  
NUMBER OF SEQ ID NOS: 282  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 2  
LENGTH: 1041  
TYPE: DNA  
ORGANISM: Streptomyces mobaraensis  
US-10-107-431-2

Query Match 81.1%; Score 14.6; DB 17; Length 1041;  
Best Local Similarity 82.4%; Pred. NO. 7.8e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTSCCGGCAT 17  
|||||:|||||  
DB 697 GACCACTCCCGGCT 713

Search completed: July 20, 2005, 22:08:23  
Job time : 466.857 secs

**THIS PAGE IS BLANK**





101	14	77.8	1701	4	US-09-949-016-2634	Sequence 2634, Ap	174	13.6	75.6	601	4	US-09-949-016-34853	Sequence 34853, A
102	14	77.8	1879	4	US-09-023-655-742	Sequence 742, App	175	13.6	75.6	601	4	US-09-949-016-34854	Sequence 34854, A
103	14	77.8	1908	1	US-08-173-508-1	Sequence 1, Appl1	176	13.6	75.6	601	4	US-09-949-016-73464	Sequence 73464, A
104	14	77.8	1908	2	US-08-265-310-1	Sequence 1, Appl1	177	13.6	75.6	601	4	US-09-949-016-73465	Sequence 73465, A
105	14	77.8	1908	3	US-08-951-742-1	Sequence 1, Appl1	178	13.6	75.6	601	4	US-09-949-016-121671	Sequence 121671, A
106	14	77.8	1920	2	US-08-855-714-1	Sequence 1, Appl1	179	13.6	75.6	601	4	US-09-949-016-135152	Sequence 135152, A
107	14	77.8	1980	4	US-09-252-991A-4190	Sequence 4190, Ap	180	13.6	75.6	601	4	US-09-949-016-135153	Sequence 135153, A
108	14	77.8	2023	3	US-09-491-522-6	Sequence 6, Appl1	181	13.6	75.6	741	4	US-09-252-991A-4070	Sequence 4070, Ap
109	14	77.8	2034	4	US-09-902-540-8514	Sequence 8514, Ap	182	13.6	75.6	885	4	US-09-902-540-5476	Sequence 5476, Ap
110	14	77.8	2070	4	US-09-252-991A-2189	Sequence 2189, Ap	183	13.6	75.6	903	4	US-09-902-540-4758	Sequence 4758, Ap
111	14	77.8	2112	4	US-09-902-540-3277	Sequence 3277, Ap	184	13.6	75.6	1164	4	US-09-252-991A-13302	Sequence 13302, A
112	14	77.8	2316	4	US-09-902-540-3869	Sequence 3869, Ap	185	13.6	75.6	1173	4	US-09-252-991A-13384	Sequence 13384, A
113	14	77.8	2450	3	US-09-491-522-2	Sequence 2, Appl1	186	13.6	75.6	1192	4	US-09-489-039A-82	Sequence 82, Appl1
114	14	77.8	2540	1	US-08-027-986-4	Sequence 4, Appl1	187	13.6	75.6	1419	4	US-09-252-991A-13796	Sequence 13796, A
115	14	77.8	2582	2	US-08-816-105A-2	Sequence 2, Appl1	188	13.6	75.6	1419	4	US-09-902-540-8278	Sequence 8278, A
116	14	77.8	3636	4	US-09-949-016-5530	Sequence 5530, Ap	189	13.6	75.6	1469	4	US-09-949-016-4983	Sequence 4983, Ap
117	14	77.8	5179	4	US-09-674-866A-1	Sequence 1, Appl1	190	13.6	75.6	1470	4	US-09-270-767-10985	Sequence 10985, A
118	14	77.8	5504	4	US-09-902-540-716	Sequence 716, App	191	13.6	75.6	1569	4	US-09-252-991A-12675	Sequence 12675, A
119	14	77.8	5883	4	US-09-949-016-5001	Sequence 5001, Ap	192	13.6	75.6	1647	4	US-09-902-540-5342	Sequence 5342, Ap
120	14	77.8	6692	3	US-09-491-522-1	Sequence 1, Appl1	193	13.6	75.6	1776	4	US-09-252-991A-4129	Sequence 4129, Ap
121	14	77.8	6855	4	US-09-902-540-897	Sequence 897, App	194	13.6	75.6	1794	4	US-09-902-540-3791	Sequence 3791, Ap
122	14	77.8	7626	4	US-09-976-594-14	Sequence 14, Appl1	195	13.6	75.6	1854	4	US-09-270-767-11990	Sequence 11900, A
123	14	77.8	7626	4	US-09-919-039-14	Sequence 14, Appl1	196	13.6	75.6	1995	4	US-09-252-991A-13546	Sequence 13546, A
124	14	77.8	7480	4	US-09-902-540-1151	Sequence 1151, Ap	197	13.6	75.6	2119	3	US-09-032-372-6	Sequence 6, Appl1
125	14	77.8	18187	4	US-09-949-016-14126	Sequence 14126, A	198	13.6	75.6	2508	4	US-09-252-991A-793	Sequence 793, App
126	14	77.8	18187	4	US-09-949-016-14127	Sequence 14127, A	199	13.6	75.6	2577	4	US-09-252-991A-828	Sequence 828, App
127	14	77.8	18700	4	US-09-949-016-13140	Sequence 13140, A	200	13.6	75.6	2730	4	US-09-489-039A-6744	Sequence 6744, Ap
128	14	77.8	19019	4	US-09-902-540-1171	Sequence 1171, Ap	201	13.6	75.6	2871	4	US-09-949-016-3998	Sequence 3998, Ap
129	14	77.8	21143	4	US-09-902-540-1191	Sequence 1191, Ap	202	13.6	75.6	2934	4	US-09-252-991A-11690	Sequence 11690, A
130	14	77.8	24905	4	US-09-902-540-1225	Sequence 1225, Ap	203	13.6	75.6	3650	4	US-09-620-312D-30	Sequence 30, Appl1
131	14	77.8	24984	4	US-09-949-016-13288	Sequence 13288, A	204	13.6	75.6	3764	4	US-09-949-016-4511	Sequence 4511, Ap
132	14	77.8	24984	4	US-09-949-016-13289	Sequence 13289, A	205	13.6	75.6	5581	4	US-09-023-655-966	Sequence 966, App
133	14	77.8	28054	4	US-09-902-540-1248	Sequence 1248, Ap	206	13.6	75.6	6604	4	US-09-949-016-16755	Sequence 16755, A
134	14	77.8	36620	4	US-09-952-060-30	Sequence 30, Appl1	207	13.6	75.6	6763	4	US-09-949-016-3399	Sequence 3399, Ap
135	14	77.8	43280	2	US-08-404-2227C-1	Sequence 1, Appl1	208	13.6	75.6	8030	4	US-09-902-540-857	Sequence 857, App
136	14	77.8	46899	1	US-08-679-279-1	Sequence 1, Appl1	209	13.6	75.6	12023	4	US-09-949-016-16253	Sequence 16253, A
137	14	77.8	47981	4	US-09-679-279-1	Sequence 1, Appl1	210	13.6	75.6	12241	4	US-09-948-138-4	Sequence 4, Appl1
138	14	77.8	49301	4	US-09-949-016-16296	Sequence 16296, A	211	13.6	75.6	13737	3	US-09-538-414-10	Sequence 10, Appl1
139	14	77.8	70563	4	US-09-949-016-16743	Sequence 16743, A	212	13.6	75.6	13737	4	US-10-074-279-10	Sequence 10, Appl1
140	14	77.8	75395	4	US-09-984-890-3	Sequence 3, Appl1	213	13.6	75.6	13857	4	US-09-620-312D-75	Sequence 75, Appl1
141	14	77.8	75395	4	US-10-274-194-3	Sequence 3, Appl1	214	13.6	75.6	14446	4	US-09-810-861B-4	Sequence 4, Appl1
142	14	77.8	91933	4	US-09-949-016-11855	Sequence 11855, A	215	13.6	75.6	14705	4	US-09-949-016-16249	Sequence 16249, A
143	14	77.8	91933	4	US-09-949-016-14628	Sequence 14628, A	216	13.6	75.6	18551	4	US-09-902-540-1187	Sequence 1187, Ap
144	14	77.8	93753	4	US-09-949-016-14573	Sequence 14573, A	217	13.6	75.6	24333	4	US-09-639-207-9	Sequence 9, Appl1
145	14	77.8	93753	4	US-09-949-016-14574	Sequence 14574, A	218	13.6	75.6	26492	4	US-09-902-540-1234	Sequence 1234, Ap
146	14	77.8	818128	4	US-09-949-016-14546	Sequence 14546, A	219	13.6	75.6	29433	4	US-09-949-016-15740	Sequence 15740, A
147	14	77.8	818128	4	US-09-949-016-14547	Sequence 14547, A	220	13.6	75.6	29899	4	US-09-902-540-1265	Sequence 1265, Ap
148	14	77.8	818128	4	US-09-949-016-14548	Sequence 14548, A	221	13.6	75.6	34534	4	US-09-949-016-15141	Sequence 15141, A
149	14	77.8	818128	4	US-09-949-016-14549	Sequence 14549, A	222	13.6	75.6	35399	4	US-09-902-540-1260	Sequence 1260, Ap
150	14	77.8	818128	4	US-09-949-016-14550	Sequence 14550, A	223	13.6	75.6	36456	4	US-09-949-016-12404	Sequence 12404, A
151	14	77.8	818128	4	US-09-949-016-14551	Sequence 14551, A	224	13.6	75.6	36457	4	US-09-949-016-13031	Sequence 13031, A
152	14	77.8	818128	4	US-09-949-016-14552	Sequence 14552, A	225	13.6	75.6	36620	4	US-09-952-060-31	Sequence 31, Appl1
153	14	77.8	818128	4	US-09-949-016-14553	Sequence 14553, A	226	13.6	75.6	36778	3	US-09-105-537-5	Sequence 5, Appl1
154	14	77.8	818128	4	US-09-949-016-14554	Sequence 14554, A	227	13.6	75.6	37215	4	US-09-949-016-15526	Sequence 15526, A
155	14	77.8	818128	4	US-09-949-016-14555	Sequence 14555, A	228	13.6	75.6	37474	4	US-09-952-060-26	Sequence 26, Appl1
156	14	77.8	818128	4	US-09-949-016-14556	Sequence 14556, A	229	13.6	75.6	38506	3	US-09-320-878-19	Sequence 19, Appl1
157	14	77.8	818128	4	US-09-949-016-14557	Sequence 14557, A	230	13.6	75.6	38506	4	US-09-141-908-1	Sequence 1, Appl1
158	14	77.8	818128	4	US-09-949-016-14558	Sequence 14558, A	231	13.6	75.6	38506	4	US-09-657-440-19	Sequence 19, Appl1
159	14	77.8	818128	4	US-09-949-016-14559	Sequence 14559, A	232	13.6	75.6	38519	4	US-09-952-060-29	Sequence 29, Appl1
160	14	77.8	818128	4	US-09-949-016-14560	Sequence 14560, A	233	13.6	75.6	88517	4	US-09-949-016-12524	Sequence 12524, A
161	14	77.8	818128	4	US-09-949-016-14561	Sequence 14561, A	234	13.6	75.6	98439	4	US-09-949-016-13557	Sequence 13557, A
162	14	77.8	818128	4	US-09-949-016-14562	Sequence 14562, A	235	13.6	75.6	134008	4	US-09-949-016-13841	Sequence 8, Appl1
163	14	77.8	818128	4	US-09-949-016-14564	Sequence 14564, A	236	13.6	75.6	154746	4	US-09-827-688-8	Sequence 8, Appl1
164	14	77.8	818128	4	US-09-949-016-14565	Sequence 14565, A	237	13.6	75.6	154746	4	US-09-827-688-8	Sequence 8, Appl1
165	14	77.8	818128	4	US-09-949-016-14566	Sequence 14566, A	238	13.6	75.6	174029	4	US-09-949-016-12610	Sequence 12610, A
166	14	77.8	818128	4	US-09-949-016-14567	Sequence 14567, A	239	13.6	75.6	174030	4	US-09-949-016-13880	Sequence 13880, A
167	14	77.8	4403765	3	US-09-103-840A-2	Sequence 2, Appl1	240	13.6	75.6	228354	4	US-09-705-400-64	Sequence 64, Appl1
168	14	77.8	4411529	3	US-09-103-840A-1	Sequence 1, Appl1	241	13.6	75.6	238815	4	US-09-949-016-16274	Sequence 16274, A
169	13.6	75.6	289	4	US-09-270-767-12817	Sequence 12817, A	242	13.6	75.6	300402	4	US-09-949-016-13632	Sequence 13632, A
170	13.6	75.6	299	4	US-09-313-294A-6210	Sequence 6210, Ap	243	13.6	75.6	784019	4	US-09-949-016-14033	Sequence 14033, A
171	13.6	75.6	376	4	US-09-621-976-8013	Sequence 8013, Ap	244	13.6	75.6	826152	4	US-09-949-016-12777	Sequence 12777, A
172	13.6	75.6	414	4	US-09-252-991A-4060	Sequence 4060, Ap	245	13.4	74.4	438	4	US-09-252-991A-15524	Sequence 15524, A
173	13.6	75.6	564	4	US-09-252-991A-4096	Sequence 4096, Ap	246	13.4	74.4	601	4	US-09-949-016-29378	Sequence 29378, A

```

C 247 13.4 74.4 601 4 US-09-949-016-29379 Sequence 29379, A
C 248 13.4 74.4 601 4 US-09-949-016-35261 Sequence 35261, A
C 249 13.4 74.4 601 4 US-09-949-016-35262 Sequence 35262, A
C 250 13.4 74.4 601 4 US-09-949-016-154544 Sequence 154544, A
C 251 13.4 74.4 601 4 US-09-949-016-154545 Sequence 154545, A
C 252 13.4 74.4 601 4 US-09-949-016-154546 Sequence 154546, A
C 253 13.4 74.4 601 4 US-09-949-016-154547 Sequence 154547, A
C 254 13.4 74.4 601 4 US-09-949-016-156823 Sequence 156823, A
C 255 13.4 74.4 601 4 US-09-949-016-156824 Sequence 156824, A
C 256 13.4 74.4 601 4 US-09-949-016-173843 Sequence 173843, A
C 257 13.4 74.4 601 4 US-09-949-016-173844 Sequence 173844, A
C 258 13.4 74.4 972 4 US-09-252-991A-6718 Sequence 6718, Ap
C 259 13.4 74.4 1212 4 US-09-252-991A-6685 Sequence 6685, Ap
C 260 13.4 74.4 1230 4 US-09-902-540-5591 Sequence 5591, Ap
C 261 13.4 74.4 1335 4 US-09-949-016-2469 Sequence 2469, Ap
C 262 13.4 74.4 1380 4 US-09-252-991A-6613 Sequence 6613, Ap
C 263 13.4 74.4 1380 4 US-09-252-991A-15862 Sequence 15862, A
C 264 13.4 74.4 1836 4 US-09-252-991A-15831 Sequence 15831, A
C 265 13.4 74.4 2055 4 US-09-252-991A-15771 Sequence 15771, A
C 266 13.4 74.4 2106 4 US-09-252-991A-15891 Sequence 15891, A
C 267 13.4 74.4 7070 4 US-09-949-016-12469 Sequence 12469, A
C 268 13.4 74.4 7070 4 US-09-949-016-15322 Sequence 15322, A
C 269 13.4 74.4 9306 3 US-09-453-702B-231 Sequence 231, App
C 270 13.4 74.4 11879 4 US-09-949-016-14211 Sequence 14211, A
C 271 13.4 74.4 14861 4 US-09-902-540-1127 Sequence 1127, Ap
C 272 13.4 74.4 15222 4 US-09-949-016-11916 Sequence 11916, A
C 273 13.4 74.4 15223 4 US-09-949-016-16912 Sequence 16912, A
C 274 13.4 74.4 18319 4 US-09-949-016-17446 Sequence 17446, A
C 275 13.4 74.4 26423 4 US-09-949-016-15459 Sequence 15459, A
C 276 13.4 74.4 38653 4 US-09-922-445-1 Sequence 1, App11
C 277 13.4 74.4 40576 4 US-09-949-016-12329 Sequence 12329, A
C 278 13.4 74.4 40577 4 US-09-949-016-16663 Sequence 16663, A
C 279 13.4 74.4 41927 4 US-09-902-540-1268 Sequence 1268, Ap
C 280 13.4 74.4 50797 4 US-09-949-016-16346 Sequence 16346, A
C 281 13.4 74.4 50797 4 US-09-949-016-16347 Sequence 16347, A
C 282 13.4 74.4 56448 4 US-09-949-016-16463 Sequence 16463, A
C 283 13.4 74.4 66627 4 US-09-949-016-12630 Sequence 12630, A
C 284 13.4 74.4 66628 4 US-09-949-016-16112 Sequence 16112, A
C 285 13.4 74.4 125672 4 US-09-949-016-16956 Sequence 16956, A
C 286 13.4 74.4 251672 4 US-09-949-016-17296 Sequence 17296, A
C 287 13.4 74.4 251682 4 US-09-949-016-11973 Sequence 11973, A
C 288 13.4 74.4 275110 4 US-09-949-016-12706 Sequence 12706, A
C 289 13.4 74.4 275110 4 US-09-949-016-16070 Sequence 16070, A
C 290 13.4 74.4 422592 4 US-09-949-016-14182 Sequence 14182, A
C 291 13.2 73.3 23 3 US-08-155-005A-15 Sequence 15, App1
C 292 13.2 73.3 23 3 US-09-363-783-15 Sequence 15, App1
C 293 13.2 73.3 23 4 US-09-661-758A-15 Sequence 15, App1
C 294 13.2 73.3 23 4 US-09-304-232-151 Sequence 151, App
C 295 13.2 73.3 239 4 US-09-016-434-463 Sequence 463, App
C 296 13.2 73.3 345 4 US-09-252-991A-15649 Sequence 15649, A
C 297 13.2 73.3 388 4 US-09-513-999C-8528 Sequence 8528, Ap
C 298 13.2 73.3 433 1 US-09-621-976-12386 Sequence 12386, A
C 299 13.2 73.3 501 1 US-08-318-193-76 Sequence 76, App1
C 300 13.2 73.3 601 4 US-09-949-016-78610 Sequence 78610, A

```

## ALIGNMENTS

```

RESULT 1
US-09-199-637A-155
Sequence 155, Application US/09199637A
Patent No. 6355411
GENERAL INFORMATION:
APPLICANT: Ausubel, Frederick
APPLICANT: Goodman, Howard M.
APPLICANT: Rahme, Laurence G.
APPLICANT: Mahajan-Miklos, Shalina
APPLICANT: Tan, Man-Wah
APPLICANT: Cao, Hui
APPLICANT: Drenkard, Eliana
APPLICANT: Tsongalis, John
TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID

```

```

TITLE OF INVENTION: SEQUENCES AND USES THEREOF
FILE REFERENCE: 00786/361002
CURRENT APPLICATION NUMBER: US/09199, 637A
CURRENT FILING DATE: 1998-11-25
PRIOR APPLICATION NUMBER: 60/066, 517
PRIOR FILING DATE: 1997-11-25
NUMBER OF SEQ ID NOS: 437
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 155
LENGTH: 513
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-155
Query Match 86.7%; Score 15.6; DB 3; Length 513;
Best Local Similarity 83.3%; Pred. No. 3.1e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Db 133 GACCACGTCCGCGCATG 150
OY 1 GACCACGTCCGCGCATG 18
|||||:|||||

```

```

RESULT 2
US-09-107-433-417/C
Sequence 417, Application US/09107433
Patent No. 6800744
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSIS
THERAPEUTICS
NUMBER OF SEQUENCES: 5206
CORRESPONDENCE ADDRESS:
ADDRESSER: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: <Unknown>
OPERATING SYSTEM: <Unknown>
SOFTWARE: <Unknown>
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107, 433
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/ 085131
FILING DATE: May 12, 1998
APPLICATION NUMBER: 60/051553
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 417:
SEQUENCE CHARACTERISTICS:
LENGTH: 612 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: misc_feature

```

## RESULT 7

## US-09-180-109A-5

; Sequence 5, Application US/09180109A  
; Patent No. 6410293  
; GENERAL INFORMATION:  
; APPLICANT: MUKUMOTO, Fujio  
; APPLICANT: NISHIO, Shoichi  
; APPLICANT: AKIMARU, Jiro  
; APPLICANT: MITSUDA, Satoshi  
; TITLE OF INVENTION: DNA Fragments Containing Biotin Biosynthetase Gene and  
; FILE REFERENCE: 0152-0490P  
; CURRENT APPLICATION NUMBER: US/09/180,109A  
; PRIOR FILING DATE: 1998-12-03  
; PRIOR APPLICATION NUMBER: 09/047838 JAPAN  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 1408  
; TYPE: DNA  
; ORGANISM: *Sphingomonas* sp.  
; FEATURE:  
; OTHER INFORMATION: Strain = SC42405  
; NAME/KEY: CDS  
; LOCATION: (202)..(1362)  
US-09-180-109A-5

Query Match 86.7%; Score 15.6; DB 3; Length 1408;  
Best Local Similarity 83.3%; Pred. No. 3e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACGTSCCGGCATG 18  
|||||:|||||

Db 1057 GACCTCGTCCGCGCATG 1074

## RESULT 8

US-09-180-109A-7  
; Sequence 7, Application US/09180109A  
; Patent No. 6410293  
; GENERAL INFORMATION:  
; APPLICANT: MUKUMOTO, Fujio  
; APPLICANT: NISHIO, Shoichi  
; APPLICANT: AKIMARU, Jiro  
; APPLICANT: MITSUDA, Satoshi  
; TITLE OF INVENTION: DNA Fragments Containing Biotin Biosynthetase Gene and  
; FILE REFERENCE: 0152-0490P  
; CURRENT APPLICATION NUMBER: US/09/180,109A  
; PRIOR FILING DATE: 1998-12-03  
; PRIOR APPLICATION NUMBER: 09/047838 JAPAN  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 1408  
; TYPE: DNA  
; ORGANISM: *Sphingomonas* sp.  
; FEATURE:  
; OTHER INFORMATION: Strain = SC42405  
; NAME/KEY: CDS  
; LOCATION: (202)..(1362)  
US-09-180-109A-7

Query Match 86.7%; Score 15.6; DB 3; Length 1408;  
Best Local Similarity 83.3%; Pred. No. 3e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACGTSCCGGCATG 18  
|||||:|||||

Db 1057 GACCTCGTCCGCGCATG 1074

## RESULT 9

US-09-199-637A-136  
; Sequence 136, Application US/09199637A  
; Patent No. 6355411  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick  
; APPLICANT: Goodman, Howard M.  
; APPLICANT: Rahme, Laurence G.  
; APPLICANT: Mahajan-Miklos, Shalina  
; APPLICANT: Tan, Man-wah  
; APPLICANT: Cao, Hui  
; APPLICANT: Drenkard, Eliana  
; APPLICANT: Teongalis, John  
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
; FILE REFERENCE: 00786/361002  
; CURRENT APPLICATION NUMBER: US/09/199,637A  
; PRIOR FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: 60/066,517  
; SOFTWARE: FastSeq for Windows Version 4.0  
; NUMBER OF SEQ ID NOS: 437  
; SEQ ID NO 136  
; LENGTH: 2048  
; TYPE: DNA  
; ORGANISM: *Pseudomonas aeruginosa*  
US-09-199-637A-136

Query Match 86.7%; Score 15.6; DB 3; Length 2048;  
Best Local Similarity 83.3%; Pred. No. 2.9e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACGTSCCGGCATG 18  
|||||:|||||

Db 1069 GACCACGTCCGCGCATG 1086

## RESULT 10

US-08-961-527-205/C  
; Sequence 205, Application US/08961527  
; Patent No. 6420135  
; GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: *Streptococcus pneumoniae* Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 391  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,527  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 205:  
; SEQUENCE CHARACTERISTICS:

LENGTH: 2395 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-961-527-205

Query Match 86.7%; Score 15.6; DB 3; Length 2395;  
Best Local Similarity 83.3%; Pred. No. 2.9e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GACCACGTSCCGGCATG 18  
|||||:|:|:|  
Db 1257 GACCACGTGCCGCATG 1240

## RESULT 11

US-09-949-016-1079  
Sequence 1079, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: C1001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1079  
LENGTH: 4927  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-1079

Query Match 86.7%; Score 15.6; DB 4; Length 4927;  
Best Local Similarity 83.3%; Pred. No. 2.7e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GACCACGTSCCGGCATG 18  
|||||:|:|:|  
Db 24 GACCACGGCCCGGCATG 41

## RESULT 12

US-09-949-016-15371/c  
Sequence 15371, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: C1001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15371  
LENGTH: 304533  
TYPE: DNA  
ORGANISM: Human  
FEATURE:

NAME/KEY: misc\_feature  
LOCATION: (1)...(304533)  
OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-15371

Query Match 86.7%; Score 15.6; DB 4; Length 304533;  
Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GACCACGTSCCGGCATG 18  
|||||:|:|:|  
Db 299865 GACCACGTGCCGCATG 299848

## RESULT 13

US-09-949-016-15372/c  
Sequence 15372, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: C1001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15372  
LENGTH: 304533  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(304533)  
OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-15372

Query Match 86.7%; Score 15.6; DB 4; Length 304533;  
Best Local Similarity 83.3%; Pred. No. 2.1e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GACCACGTSCCGGCATG 18  
|||||:|:~|:|:|  
Db 299865 GACCACGTGCCGCATG 299848

## RESULT 14

US-09-949-016-17330  
Sequence 17330, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
FILE REFERENCE: C1001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
PRIOR FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 17330  
LENGTH: 10948

TYPE: DNA  
ORGANISM: Human  
US-09-949-016-17310

Query Match 83.3%; Score 15; DB 4; Length 10948;  
Best Local Similarity 88.2%; Pred. No. 5e+02;  
Matches 15; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ACCACGTSCCGGCATG 18  
|||||:|||||  
Db 4160 ACCACGTCCCTCGCATG 4176

RESULT 15  
US-09-621-976-16915/c  
Sequence 16915, Application US/09621976  
Patent No. 6639063  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, J.B.  
APPLICANT: Jobert, S.  
TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
FILE REFERENCE: GENSET 054PR2  
CURRENT APPLICATION NUMBER: US/09/621,976  
CURRENT FILING DATE: 2000-07-21  
NUMBER OF SEQ ID NOS: 19335  
SOFTWARE: Patent.pm  
SEQ ID NO 16915  
LENGTH: 510  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-621-976-16915

Query Match 81.1%; Score 14.6; DB 4; Length 510;  
Best Local Similarity 82.4%; Pred. No. 9.1e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACGTSCCGGCAT 17  
|||||:|||||  
Db 59 GACCACGTGCCGTCAT 43

RESULT 16  
US-09-640-211A-238/c  
Sequence 238, Application US/09640211A  
Patent No. 6833446  
GENERAL INFORMATION:  
APPLICANT: Wood, Marion  
APPLICANT: Shenk, Michael A.  
APPLICANT: McGrath, Annette  
APPLICANT: Glenn, Matthew  
TITLE OF INVENTION: Compositions and Methods for the  
FILE REFERENCE: 11000.1021CIU  
CURRENT APPLICATION NUMBER: US/09/640,211A  
CURRENT FILING DATE: 2000-08-16  
NUMBER OF SEQ ID NOS: 2368  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 238  
LENGTH: 521  
TYPE: DNA  
ORGANISM: Eucalyptus grandis  
US-09-640-211A-238

Query Match 81.1%; Score 14.6; DB 4; Length 521;  
Best Local Similarity 82.4%; Pred. No. 9.1e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACGTSCCGGCAT 17  
|||||:|||||  
Db 497 GACCACGTCCCGCGCT 481

RESULT 17  
US-09-640-211A-1210/c  
Sequence 1210, Application US/09640211A  
Patent No. 6833446  
GENERAL INFORMATION:  
APPLICANT: Wood, Marion  
APPLICANT: Shenk, Michael A.  
APPLICANT: McGrath, Annette  
APPLICANT: Glenn, Matthew  
TITLE OF INVENTION: Compositions and Methods for the  
FILE REFERENCE: 11000.1021CIU  
CURRENT APPLICATION NUMBER: US/09/640,211A  
CURRENT FILING DATE: 2000-08-16  
NUMBER OF SEQ ID NOS: 2368  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1210  
LENGTH: 521  
TYPE: DNA  
ORGANISM: Eucalyptus grandis  
US-09-640-211A-1210

Query Match 81.1%; Score 14.6; DB 4; Length 521;  
Best Local Similarity 82.4%; Pred. No. 9.1e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACGTSCCGGCAT 17  
|||||:|||||  
Db 497 GACCACGTCCCGCGCT 481

RESULT 18  
US-09-902-540-7926  
Sequence 7926, Application US/09902540  
Patent No. 6833447  
GENERAL INFORMATION:  
APPLICANT: Goldman, Barry S.  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Wiegand, Roger C.  
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(15849)B  
CURRENT APPLICATION NUMBER: US/09/902,540  
CURRENT FILING DATE: 2001-07-10  
PRIOR APPLICATION NUMBER: 60/217,883  
PRIOR FILING DATE: 2000-07-10  
NUMBER OF SEQ ID NOS: 16825  
SEQ ID NO 7926  
LENGTH: 669  
TYPE: DNA  
ORGANISM: Myxococcus xanthus  
US-09-902-540-7926

Query Match 81.1%; Score 14.6; DB 4; Length 669;  
Best Local Similarity 82.4%; Pred. No. 8.9e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ACCACGTSCCGGCATG 18  
|||||:|||||  
Db 497 ACCACGTACCGGCATG 513

RESULT 19  
US-08-831-132-13/c  
Sequence 13, Application US/08831132  
Patent No. 6008322  
GENERAL INFORMATION:  
APPLICANT: Kuestner, Rolf E.  
APPLICANT: Conklin, Darrell C.  
APPLICANT: Lok, Si  
APPLICANT: Buddle, Michele  
APPLICANT: Downey, William  
TITLE OF INVENTION: STANNICALCIN-2

```

;
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zymogenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA: Patent Release #1.0, Version #1.25
; APPLICATION NUMBER: US/08/831,132
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sawislak, Deborah A.
; REGISTRATION NUMBER: 37,438
; REFERENCE/DOCKET NUMBER: 96-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6672
; TELEFAX: 206-442-6672
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 888 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 1..72
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 73..888
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..888
; US-08-831-132-13
;
; Query Match 81.1%; Score 14.6; DB 3; Length 888;
; Best Local Similarity 82.4%; Pred. No. 8.8e+02;
; Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 GACCACGTCGCGGCAT 17
; DB 192 GCCCAGTCGCCGCGCAT 176
;
; RESULT 20
; US-09-416-150-13/c
; Sequence 13, Application US/09416150
; Patent No. 6171822
; GENERAL INFORMATION:
; APPLICANT: Kuestner, Rolf E.
; Konklin, Darrell C.
; Lok, Si
; Buddle, Michele
; Downey, William
; TITLE OF INVENTION: STANNIOLALCIN-2
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zymogenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/416,150
; FILING DATE: 11-Oct-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/831,132
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Sawislak, Deborah A.
; REGISTRATION NUMBER: 37,438
; REFERENCE/DOCKET NUMBER: 96-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6672
; TELEFAX: 206-442-6672
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 888 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 1..72
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 73..888
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..888
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
;
; US-09-416-150-13
;
; Query Match 81.1%; Score 14.6; DB 3; Length 888;
; Best Local Similarity 82.4%; Pred. No. 8.8e+02;
; Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 GACCACGTCGCGGCAT 17
; DB 192 GCCCAGTCGCCGCGCAT 176
;
; RESULT 21
; US-09-902-540-7901/c
; Sequence 7901, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(11549)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 7901
; LENGTH: 1173
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; US-09-902-540-7901
;
; Query Match 81.1%; Score 14.6; DB 4; Length 1173;
; Best Local Similarity 82.4%; Pred. No. 8.6e+02;
; Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 GACCACGTCGCGGCAT 17
; DB 684 GACCACGTCGCCGCGCT 668

```



```
RESULT 22
US-09-902-540-7980/c
; Sequence 7980, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 7980
; LENGTH: 1347
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-7980

Query Match      81.1%; Score 14.6; DB 4; Length 1347;
Best Local Similarity 82.4%; Pred. No. 8.6e+02;
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 GACCACGTSCCGGCAT 17
Db      1149 GACCACGTGCGCGGCAT 1133

RESULT 23
US-09-270-767-15069
; Sequence 15069, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-034
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15069
; LENGTH: 1424
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-15069

Query Match      81.1%; Score 14.6; DB 4; Length 1424;
Best Local Similarity 82.4%; Pred. No. 8.6e+02;
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 GACCACGTSCCGGCAT 17
Db      751 GTCCACGTGCGCGGCAT 767

RESULT 24
US-09-821-167-13
; Sequence 13, Application US/09821167
; Patent No. 656968
; GENERAL INFORMATION:
; APPLICANT: Hosnan, Ann C.
; TITLE OF INVENTION: Isolated Nucleic Acids from Micromonospora rosaria
; FILE REFERENCE: INO11490
; CURRENT APPLICATION NUMBER: US/09/821,167
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: US 60/194,461
```

```
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1749
; TYPE: DNA
; ORGANISM: Micromonospora rosaria
US-09-821-167-13

Query Match      81.1%; Score 14.6; DB 4; Length 1749;
Best Local Similarity 82.4%; Pred. No. 8.5e+02;
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 GACCACGTSCCGGCAT 17
Db      436 GACCACGTGCGCGGCAT 452

RESULT 25
US-09-949-016-2932
; Sequence 2932, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2932
; LENGTH: 3966
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-2932

Query Match      81.1%; Score 14.6; DB 4; Length 3966;
Best Local Similarity 82.4%; Pred. No. 8.1e+02;
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      2 ACCACGTSCCGGCATG 18
Db      2385 ACCACGTGCGCGGCATG 2401

RESULT 26
US-09-949-016-238
; Sequence 238, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 238
; LENGTH: 4979
```

```

; TYPE: DNA
; ORGANISM: Human
US-09-949-016-238

Query Match
Best Local Similarity 81.1%; Score 14.6; DB 4; Length 4979;
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACCACGTSCCGGCATG 18
    |||||:|:|:|
    4484 ACCACGTSCCGGCATG 4500

RESULT 27
US-09-902-540-808
; Sequence 808, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 808
; LENGTH: 6012
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-808

Query Match
Best Local Similarity 81.1%; Score 14.6; DB 4; Length 6012;
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTSCCGGCAT 17
    |||||:|:|:|
    1922 GACCACTSCCGGCAT 1938

RESULT 28
US-09-902-540-799/c
; Sequence 799, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 799
; LENGTH: 6488
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-799

Query Match
Best Local Similarity 81.1%; Score 14.6; DB 4; Length 6488;
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 ACCACGTSCCGGCATG 18
    |||||:|:|:|
    3302 ACCACGTSCCGGCATG 3286
```

```

; TYPE: DNA
; ORGANISM: Human
US-09-902-540-796

Query Match
Best Local Similarity 81.1%; Score 14.6; DB 4; Length 6603;
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTSCCGGCAT 17
    |||||:|:|:|
    5510 GACCACTSCCGGCAT 5526

RESULT 29
US-09-902-540-796
; Sequence 796, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 796
; LENGTH: 6603
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-796

Query Match
Best Local Similarity 81.1%; Score 14.6; DB 4; Length 6603;
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTSCCGGCAT 17
    |||||:|:~|:|
    5510 GACCACTSCCGGCAT 5526

RESULT 30
US-09-028-934-28
; Sequence 28, Application US/09028934
; Patent No. 6117670
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Lam, Steven T.
; APPLICANT: Hammer, Philip E.
; APPLICANT: van Pee, Karl-Heinz
; APPLICANT: Kimer, Sabine
; APPLICANT: Young, Thomas R.
; TITLE OF INVENTION: Pyrolytic Biosynthesis Genes and Uses
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 6117670artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/028,934
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
```

REFERENCE/DOCKET NUMBER: CGC1506/CIP7  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8931 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Burkholderia cepacia  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 657..2267  
OTHER INFORMATION: /product= "PrnA"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2270..3355  
OTHER INFORMATION: /product= "PrnB"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3421..5121  
OTHER INFORMATION: /product= "PrnC"  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 5145..6266  
OTHER INFORMATION: /product= "PrnD"  
US-09-028-934-28

Query Match 81.1%; Score 14.6; DB 3; Length 8931;  
Best Local Similarity 82.4%; Pred. No. 7.7e+02;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GACGACGTSCCGGCAT 17  
|||:|||||  
Db 2387 GACGACGTSCCGGCAT 2403

Search completed: July 20, 2005, 16:48:19  
Job time: 115.286 secs

**THIS PAGE IS BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: July 20, 2005, 15:46:36 ; Search time 2237.57 Seconds

(without alignments)  
306.206 Million cell updates/sec

Title: US-10-017-471b-1

Perfect score: 18  
Sequence: 1 gaccacgtccggcgcacg 18Scoring table: IDENTITY NUC  
Gapop 10'-0, Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 300 summaries

## Database :

EST: \*  
1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_est3: \*  
4: gb\_est4: \*  
5: gb\_est5: \*  
6: gb\_est6: \*  
7: gb\_est7: \*  
8: gb\_est8: \*  
9: gb\_est9: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	17.2	95.6	742	9	CR810834
C 2	16.2	90.0	376	4	BI634046
C 3	16.2	90.0	703	6	CB016461
C 4	16.2	90.0	738	7	CN225729
C 5	15.6	86.7	223	2	BF813258
C 6	15.6	86.7	293	2	BB397065
C 7	15.6	86.7	293	6	CB995647
C 8	15.6	86.7	433	2	BF704768
C 9	15.6	86.7	435	2	BF656537
C 10	15.6	86.7	451	2	BB729006
C 11	15.6	86.7	510	8	CL797916
C 12	15.6	86.7	514	8	AQ397255
C 13	15.6	86.7	523	7	CO344004
C 14	15.6	86.7	539	1	AL903319
C 15	15.6	86.7	564	9	LBAPB3C09
C 16	15.6	86.7	581	5	BP221909
C 17	15.6	86.7	582	5	BP356698
C 18	15.6	86.7	583	5	BP280394
C 19	15.6	86.7	583	5	BP359091
C 20	15.6	86.7	587	5	BP364824
C 21	15.6	86.7	588	5	BP310516
C 22	15.6	86.7	588	8	AZ411523
C 23	15.6	86.7	602	1	AI389668
C 24	15.6	86.7	602	1	AI389668

C 25	15.6	86.7	612	6	CN721022
C 26	15.6	86.7	624	7	CN720685
C 27	15.6	86.7	631	6	CA100015
C 28	15.6	86.7	634	6	CA071958
C 29	15.6	86.7	636	7	CF367459
C 30	15.6	86.7	641	6	CB096877
C 31	15.6	86.7	655	6	CD629831
C 32	15.6	86.7	664	9	CL786031
C 33	15.6	86.7	666	6	CD629825
C 34	15.6	86.7	678	5	AG130601
C 35	15.6	86.7	701	6	CD629828
C 36	15.6	86.7	702	6	CD629833
C 37	15.6	86.7	713	6	CD629839
C 38	15.6	86.7	715	6	CD629829
C 39	15.6	86.7	727	9	CL720198
C 40	15.6	86.7	744	6	CD629841
C 41	15.6	86.7	744	9	CG675031
C 42	15.6	86.7	766	8	AQ698996
C 43	15.6	86.7	795	5	CB997742
C 44	15.6	86.7	799	5	BX616161
C 45	15.6	86.7	804	2	AW940798
C 46	15.6	86.7	853	8	BZ553300
C 47	15.6	86.7	992	5	BQ719203
C 48	15.6	86.7	999	5	BQ188259
C 49	15.6	86.7	1000	9	CNS01086
C 50	15.6	86.7	1030	5	BM903853
C 51	15.6	86.7	1351	3	CR724745
C 52	15.6	86.7	1557	3	CR728715
C 53	15.6	86.7	1560	3	CR726317
C 54	15.6	86.7	1601	3	CR702207
C 55	15.6	86.7	1608	3	CR727784
C 56	15.6	86.7	1622	3	CR730580
C 57	15.6	86.7	1634	3	CR723875
C 58	15.6	86.7	1655	3	CR733878
C 59	15.2	84.4	359	8	AZ468435
C 60	15.2	84.4	467	8	AZ496471
C 61	15.2	84.4	552	2	BF077253
C 62	15.2	84.4	561	2	BE236993
C 63	15.2	84.4	566	2	BE237830
C 64	15.2	84.4	611	2	BF490349
C 65	15.2	84.4	686	9	CR027543
C 66	15.2	84.4	709	8	BH950057
C 67	15.2	84.4	714	4	BJ735213
C 68	15.2	84.4	719	4	BJ745844
C 69	15.2	84.4	738	2	BF627492
C 70	15.2	84.4	777	8	BZ065629
C 71	15.2	84.4	819	8	BZ220159
C 72	15.2	84.4	849	9	CG100102
C 73	15.2	84.4	917	9	CG100102
C 74	15.2	84.4	937	9	CG100099
C 75	15.2	84.4	1195	5	BM907019
C 76	15.2	84.4	184	4	BM484770
C 77	15.2	84.4	246	6	AZ888642
C 78	15.2	84.4	271	6	CA779153
C 79	15.2	84.4	297	7	CN029046
C 80	15.2	84.4	317	4	BI344931
C 81	15.2	84.4	317	4	BI345070
C 82	15.2	84.4	395	4	BI118321
C 83	15.2	84.4	409	6	CB547063
C 84	15.2	84.4	412	1	AA609648
C 85	15.2	84.4	482	7	CF769783
C 86	15.2	84.4	496	5	EX926321
C 87	15.2	84.4	500	5	EX915300
C 88	15.2	84.4	546	9	CE122049
C 89	15.2	84.4	553	9	CE126154
C 90	15.2	84.4	564	6	CB286775
C 91	15.2	84.4	651	6	CL602170
C 92	15.2	84.4	672	2	BM539336
C 93	15.2	84.4	721	7	CK460965
C 94	15.2	84.4	770	5	BD108056
C 95	15.2	84.4	791	6	DX779315
C 96	15.2	84.4	837	9	CC494591
C 97	15.2	84.4	883	7	CO165983

C 98	15	83.3	926	6	CD791885	EST63246	C 171	14.6	81.1	377	5	BY011149
C 99	15	83.3	936	6	CD779314	EST650675	C 172	14.6	81.1	377	5	BY017349
C 100	15	83.3	991	6	CNS0248B	AK053442	C 173	14.6	81.1	377	5	BY205866
C 101	15	83.3	3612	3	AK053442	Mus_muscul	C 174	14.6	81.1	378	5	BY015000
C 102	14.8	82.2	914	2	AL574513	CC178278	C 175	14.6	81.1	378	5	BY016295
C 103	14.6	81.1	147	8	CC178278	AK053442	C 176	14.6	81.1	379	5	BY118960
C 104	14.6	81.1	170	7	CK098978	AK053442	C 177	14.6	81.1	380	5	BY23592
C 105	14.6	81.1	178	5	BQ290364	h045P68.5	C 178	14.6	81.1	381	5	BY229934
C 106	14.6	81.1	182	6	CA568095	K0423FP01	C 179	14.6	81.1	381	7	CO946683
C 107	14.6	81.1	182	9	CE711187	t1gr-g88	C 180	14.6	81.1	383	5	BY013340
C 108	14.6	81.1	234	9	CE566457	t1gr-g88	C 181	14.6	81.1	383	5	BY068832
C 109	14.6	81.1	238	1	AA390339	UD03283.5	C 182	14.6	81.1	383	5	BY175209
C 110	14.6	81.1	264	2	AW815436	QV0-ST021	C 183	14.6	81.1	385	2	BF149921
C 111	14.6	81.1	278	8	BZ477857	BONON49TF	C 184	14.6	81.1	386	2	BY027567
C 112	14.6	81.1	280	2	AM454211	zeH11101	C 185	14.6	81.1	387	2	BE439307
C 113	14.6	81.1	291	5	BY281797	BY281797	C 186	14.6	81.1	387	5	BY235517
C 114	14.6	81.1	294	2	AM428473	66561 MAR	C 187	14.6	81.1	389	5	BY267409
C 115	14.6	81.1	295	6	CD553835	B0366D12	C 188	14.6	81.1	391	2	BB842330
C 116	14.6	81.1	298	5	BP138456	BP138456	C 189	14.6	81.1	393	5	BB842330
C 117	14.6	81.1	307	5	BY333719	BY333719	C 190	14.6	81.1	394	5	BY278370
C 118	14.6	81.1	310	5	BE911620	601663302	C 191	14.6	81.1	400	5	CP644392
C 119	14.6	81.1	316	2	BE911620	601663302	C 192	14.6	81.1	400	5	CP644392
C 120	14.6	81.1	327	6	BY782882	BY782882	C 193	14.6	81.1	400	6	CP699634
C 121	14.6	81.1	329	5	BY181592	BY181592	C 194	14.6	81.1	401	4	BT166039
C 122	14.6	81.1	329	5	BY191027	BY191027	C 195	14.6	81.1	404	5	BY131409
C 123	14.6	81.1	329	6	BY789463	BY789463	C 196	14.6	81.1	405	1	AL168832
C 124	14.6	81.1	332	5	BY325473	BY325473	C 197	14.6	81.1	408	5	BY264392
C 125	14.6	81.1	333	5	BY324375	BY324375	C 198	14.6	81.1	409	7	BY019283
C 126	14.6	81.1	334	5	BY073543	BY073543	C 199	14.6	81.1	409	5	CP651819
C 127	14.6	81.1	334	5	BY329302	BY329302	C 200	14.6	81.1	410	7	CP651819
C 128	14.6	81.1	340	5	BY189875	BY189875	C 201	14.6	81.1	414	6	CL19902
C 129	14.6	81.1	342	5	BY010190	BY010190	C 202	14.6	81.1	415	4	BM087239
C 130	14.6	81.1	342	5	BY213993	BY213993	C 203	14.6	81.1	417	5	BY273289
C 131	14.6	81.1	343	5	BY321572	BY321572	C 204	14.6	81.1	418	7	CL213467
C 132	14.6	81.1	344	2	BF511023	UI-H-B14	C 205	14.6	81.1	418	9	CL213467
C 133	14.6	81.1	347	5	BY322040	BY322040	C 206	14.6	81.1	419	2	BB850083
C 134	14.6	81.1	348	1	AI089097	o81Hn12.8	C 207	14.6	81.1	419	4	BT668730
C 135	14.6	81.1	349	5	BY044263	BY044263	C 208	14.6	81.1	419	5	CL586760
C 136	14.6	81.1	350	5	BY044263	BY044263	C 209	14.6	81.1	420	5	BY275542
C 137	14.6	81.1	350	5	BY162806	BY162806	C 210	14.6	81.1	420	7	BY223312
C 138	14.6	81.1	350	5	BY335693	BY335693	C 211	14.6	81.1	423	5	BY223312
C 139	14.6	81.1	352	5	BY163418	BY163418	C 212	14.6	81.1	424	2	BB848780
C 140	14.6	81.1	353	5	BY043994	BY043994	C 213	14.6	81.1	424	5	BY281952
C 141	14.6	81.1	354	5	BY231635	BY231635	C 214	14.6	81.1	424	5	BY455103
C 142	14.6	81.1	355	1	AA735789	GM09866.5	C 215	14.6	81.1	425	5	BY259533
C 143	14.6	81.1	355	5	BY324093	BY324093	C 216	14.6	81.1	426	2	BB850960
C 144	14.6	81.1	356	5	BY023197	BY023197	C 217	14.6	81.1	426	5	BY295370
C 145	14.6	81.1	356	5	BY044034	BY044034	C 218	14.6	81.1	426	6	CB794102
C 146	14.6	81.1	356	5	BY066574	BY066574	C 219	14.6	81.1	430	2	BB849148
C 147	14.6	81.1	356	5	BY182140	BY182140	C 220	14.6	81.1	431	5	BY241970
C 148	14.6	81.1	356	5	BY223601	BY223601	C 221	14.6	81.1	431	5	BY250665
C 149	14.6	81.1	358	5	BY071958	BY071958	C 222	14.6	81.1	431	7	CN819440
C 150	14.6	81.1	358	5	BY181989	BY181989	C 223	14.6	81.1	433	2	BB849589
C 151	14.6	81.1	358	5	BY203119	BY203119	C 224	14.6	81.1	433	5	BP087477
C 152	14.6	81.1	359	5	BY026243	BY026243	C 225	14.6	81.1	434	6	CA620032
C 153	14.6	81.1	359	5	BY231039	BY231039	C 226	14.6	81.1	435	2	AW358700
C 154	14.6	81.1	360	5	BY185924	BY185924	C 227	14.6	81.1	443	7	CF681873
C 155	14.6	81.1	360	5	BY187856	BY187856	C 228	14.6	81.1	443	7	CN819932
C 156	14.6	81.1	361	5	BY009542	BY009542	C 229	14.6	81.1	444	8	CB788315
C 157	14.6	81.1	361	5	BY013030	BY013030	C 230	14.6	81.1	444	8	BZ293004
C 158	14.6	81.1	362	5	BY146507	BY146507	C 231	14.6	81.1	446	5	BY271817
C 159	14.6	81.1	362	5	BY176154	BY176154	C 232	14.6	81.1	446	9	BX907061
C 160	14.6	81.1	365	2	BF162208	601770688	C 233	14.6	81.1	447	7	RI6064
C 161	14.6	81.1	365	5	BY088058	BY088058	C 234	14.6	81.1	452	7	CO284307
C 162	14.6	81.1	366	5	BY022237	BY022237	C 235	14.6	81.1	457	4	BS960052
C 163	14.6	81.1	368	5	BY235346	BY235346	C 236	14.6	81.1	457	6	CA542855
C 164	14.6	81.1	371	2	BB844425	BB844425	C 237	14.6	81.1	458	7	AA803236
C 165	14.6	81.1	371	5	BY012546	BY012546	C 238	14.6	81.1	460	1	CF896385
C 166	14.6	81.1	371	5	BY230476	BY230476	C 239	14.6	81.1	460	1	AA817646
C 167	14.6	81.1	372	5	BY014420	BY014420	C 240	14.6	81.1	467	2	BB857603
C 168	14.6	81.1	372	5	BY169527	BY169527	C 241	14.6	81.1	468	2	BB855178
C 169	14.6	81.1	374	2	BB842553	BB842553	C 242	14.6	81.1	470	2	BF751452
C 170	14.6	81.1	376	5	BY028119	BY028119	C 243	14.6	81.1	473	1	AA942179
												BY011149
												BY017349
												BY205866
												BY015000
												BY016295
												BY118960
												BY23592
												BY229934
												CO946683
												BY013340
												BY068832
												BY175209
												BF149921
												UY07567
												CD01092
												BE439307
												BY235317
												BY267409
												BY267409
												BB842330
												BY278370
												BY267085
												TYEST-A05
												BY133265
												AMGNNUC
												RE06059.5
												BY131409
												BY131409
												LY76612.X
												BY264392
												BY019283
												PMCC02XD
												21522
												CD4-1
												CL19902
												500036
												BY273289
												HRO4408.H
												BY223312
												BB848780
												BY281952
												BY455103
												BY259533
												BB850960
												AMGNNUC
												BB849148
												BY241970
												BY250665
												HRO4406.F
												BB849589
												AMGNNUC
												CH240
												BY271817
												Leishman
												Y449B01.i1

C 244	14.6	81.1	473	1	AT700719	w39g02.x
C 245	14.6	81.1	473	2	BE021973	BB021973 em64g1v.y
C 246	14.6	81.1	474	2	BB863466	BB863466
C 247	14.6	81.1	476	5	BX513312	BX513312 BX513312
C 248	14.6	81.1	479	2	BF765777	BF765777 287414.MA
C 249	14.6	81.1	483	7	CO298673	CO298673 EX173457
C 250	14.6	81.1	486	4	BB633856	BB633856 AT29806.5
C 251	14.6	81.1	487	2	BF89536	BF89536 289067.MA
C 252	14.6	81.1	488	6	CB726154	CB726154 AMGNMUC:M
C 253	14.6	81.1	488	6	CG667743	CG667743 OST461947
C 254	14.6	81.1	490	7	CO719318	CO719318 DG14-8101
C 255	14.6	81.1	493	8	AZ853001	AZ853001 2M0155B14
C 256	14.6	81.1	500	9	CG540795	CG540795 OST133244
C 257	14.6	81.1	503	1	AA423593	AA423593 ve76f01.x
C 258	14.6	81.1	507	7	AA423593	AA423593 ve76f01.x
C 259	14.6	81.1	508	7	CG436786	CG436786 BE04010A2
C 260	14.6	81.1	509	9	CE720364	CE720364 tigr-g88-
C 261	14.6	81.1	509	9	CL556189	CL556189 OB-BAC00
C 262	14.6	81.1	510	8	BB860156	BB860156 A3523-1T7
C 263	14.6	81.1	512	1	AA438361	AA438361 LD12621.5
C 264	14.6	81.1	514	6	CD548552	CD548552 B0291E08-
C 265	14.6	81.1	517	1	A1390058	A1390058 GH21527.3
C 266	14.6	81.1	517	4	BT340188	BT340188 365382.MA
C 267	14.6	81.1	520	1	AV390930	AV390930
C 268	14.6	81.1	521	1	BI641964	BI641964 SD25653.5
C 269	14.6	81.1	522	2	BB867564	BB867564
C 270	14.6	81.1	522	3	CNS09676	CNS09676
C 271	14.6	81.1	523	8	AZ583137	AZ583137 1M0376G19
C 272	14.6	81.1	524	7	CF572824	CF572824 MCSA062A0
C 273	14.6	81.1	526	1	A1518223	A1518223 LD37676.5
C 274	14.6	81.1	529	4	BM176280	BM176280 TGBSTZYB3
C 275	14.6	81.1	529	7	CF862838	CF862838 p8ZS005XA
C 276	14.6	81.1	530	4	BI375435	BI375435 RE63251.5
C 277	14.6	81.1	530	4	CF863617	CF863617 p8ZS007XL
C 278	14.6	81.1	538	7	CF862232	CF862232 p8ZS002XO
C 279	14.6	81.1	538	7	CF867985	CF867985 E0130C07-
C 280	14.6	81.1	540	1	A1531270	A1531270 SD02237.5
C 281	14.6	81.1	540	6	CA717718	CA717718 wdr4c.pk0
C 282	14.6	81.1	542	2	AM213026	AM213026 un98d08.Y
C 283	14.6	81.1	542	5	B0551605	B0551605 H4010D04
C 284	14.6	81.1	545	1	A1258895	A1258895 LP02194.5
C 285	14.6	81.1	545	1	BI228547	BI228547 RE26115.5
C 286	14.6	81.1	546	7	A1549761	A1549761 ve76f01.Y
C 287	14.6	81.1	546	7	CO156796	CO156796 EN09326.5
C 288	14.6	81.1	551	2	AM943954	AM943954 LD48009.3
C 289	14.6	81.1	551	4	BI640684	BI640684 SD23653.5
C 290	14.6	81.1	552	7	CF246273	CF246273 TGBSTZY16
C 291	14.6	81.1	553	4	BI168680	BI168680 RE09524.5
C 292	14.6	81.1	558	7	CF945234	CF945234 T1EST-A12
C 293	14.6	81.1	560	7	CK737916	CK737916 TGBSTZY18
C 294	14.6	81.1	562	2	BE144120	BE144120 MR0-HT016
C 295	14.6	81.1	562	7	CF247768	CF247768 TGBSTZY14
C 296	14.6	81.1	563	8	BI2394071	BI2394071 E1NBP32TR
C 297	14.6	81.1	568	4	BI237871	BI237871 RE34190.5
C 298	14.6	81.1	569	4	BG974491	BG974491 602844283
C 299	14.6	81.1	572	4	BG578433	BG578433 de99a06.Y
C 300	14.6	81.1	573	7	CF863689	CF863689 p8ZS007XO

## ALIGNMENTS

RESULT 1  
CR810834/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

CR810834  
G00AA34BH01FM1 INRA BAC Bos taurus genomic clone INRAb\_566B08, DNA  
sequence, genomic survey sequence.  
CR810834  
CR810834.1 GI:52692005  
GSS.  
Bos taurus (cow)  
Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

742 bp  
DNA  
linear  
GSS 24-SEP-2004

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 742)  
Eggen, A., Schibler, L. and Roy, A.  
Bovine BAC End Sequences from the INRA bovine BAC library  
Unpublished  
2 (bases 1 to 742)  
Genoscope.  
Direct Submission  
Submitted (20-SEP-2004) Genoscope - Centre National de Sequencage :  
BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)  
- Web : www.genoscope.cns.fr  
Contact: Andre Eggen  
Department of Animal Genetics - IGBC  
INRA  
78350 Jouy-en-Josas, France  
Tel: 33 1 34 65 24 24  
Fax: 33 1 34 65 24 78  
Email: eggen@jouy.inra.fr  
Emails are derived from the INRA bovine BAC library  
Closes are derived from the INRA bovine BAC library  
(http://locus.jouy.inra.fr/fpc/cattle.bac.map.htm). For BAC library  
availability, please contact Andre Eggen (eggen@jouy.inra.fr). This  
work was undertaken as part of the International Bovine BAC  
Mapping Consortium (IBBMC) by INRA (Jouy-en-Josas) and Genoscope  
(Evry) Plate: 586 row: B column: 08  
Seq primer: M13 forward  
Class: BAC ends.

FEATURES  
source  
location/Qualifiers  
1..742  
/organism="Bos taurus"  
/mol\_type="genomic DNA"  
/strain="Breed: Holstein"  
/db\_xref="taxon:9913"  
/clone="INRA\_566B08"  
/sex="Male"  
/cell\_type="fibroblast"  
/clone\_1ib="INRA bovine BAC"  
/note="Vector: pBeloBAC11; Site 1: HindIII; Holstein bull;  
INRA Bovine BAC library (Male) produced by Andre  
Eggen-Genoscope sequence ID : G00AA34BH01FM1"

ORIGIN  
Query Match  
Best local Similarity 88.9%; Pred. No. 6.5e+02;  
Matches 16; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY  
Db  
1 GACCACGTSCCGGCATG 18  
287 GACCACGTSCCGGCATG 270

RESULT 2  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

BI634046  
376 bp  
mRNA  
linear  
EST 10-SEP-2001  
SD28283, Spittle SD Drosophila melanogaster Schneider I2 cell culture  
PORT2 Drosophila melanogaster cDNA clone SD28283 5 similar to  
Pro626: FBan0004097 GO:[208 core proteasome (GO:0005839);  
multicatalytic endopeptidase (GO:0004299); 208 core proteasome  
(GO:0005839); multicatalytic endopeptidase (GO:0004299)] located  
on: 3L 73A10-73A10; 05/24/2001, mRNA sequence.  
BI634046  
BI634046.1 GI:15536256  
EST.  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 376)  
Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G.,  
Lewis, S. and Rubin, G.M.  
BDGP/HMMI Drosophila EST Project  
Unpublished (2001)

COMMENT Contact: Stapleton, M.  
BDGP  
Lawrence Berkeley National Lab  
One Cyclotron Rd, Berkeley, CA 94720, USA  
Fax: 510 486 6798  
Email: <http://www.fruitfly.org/EST/estfruitfly.berkeley.edu>  
hit genomic AB003526: arm:3L [16482650..16763536]  
estimated-cyto:73A8-73D1: 05/24/2001  
Plate: SD.282 row: G Column: 11  
High quality sequence stop: 276.  
Location/Qualifiers  
1..376  
/organism="Drosophila melanogaster"  
/mol\_type="mRNA"  
/db\_xref="taxon:7227"  
/clone="SD28283"  
/lab\_host="DHS-alpha"  
/clone\_lib="SD Drosophila melanogaster Schneider L2 cell  
culture POT2"  
/note="Vector: POT2; Site\_1: EcoRI; Site\_2: XhoI; Sized  
fractionated cDNAs were directly ligated into POT2.  
Plasmid cDNA library."

ORIGIN  
Query Match 90.0%; Score 16.2; DB 4; Length 376;  
Best Local Similarity 88.2%; Pred. No. 2e+03; Indels 0; Gaps 0;  
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACCAGTSCCGGCATG 18  
|||||:|||||  
Db 101 ACCAGTCCCGGCATG 117

RESULT 3  
CB016461 703 bp mRNA linear EST 10-JAN-2003  
LOCUS pgnic.pk012.h12 Chicken lymphoid cDNA library (pgnic) Gallus gallus  
DEFINITION cDNA clone pgnic.pk012.h12 5' similar to gbl/AM2739.1 similar to  
prion protein interacting protein [Mus musculus], mRNA sequence.  
ACCESSION CB016461  
VERSION CB016461.1 GI:27591197  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 703)  
Morgan, R.W. and Burnside, J.  
Chicken ESTs from lymphoid tissue  
Unpublished (2002)  
Contact: Robin W. Morgan  
University of Delaware  
Townsend Hall, Newark, DE 19717, USA  
Tel: 302-831-1341  
Fax: 302-831-2822  
Email: [morganudel.edu](mailto:morganudel.edu), [www.chickest.udel.edu](http://www.chickest.udel.edu).

FEATURES  
source  
1..703  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9031"  
/clone="pgnic.pk012.h12"  
/sex="Male and Female"  
/tissue\_type="thymus, bursa, spleen, PBL, bone marrow"  
/lab\_host="E.coli EMD108"  
/clone\_lib="Chicken lymphoid cDNA library (pgnic)"  
/note="Vector: PCWVSPORT 6"

ORIGIN  
Query Match 90.0%; Score 16.2; DB 6; Length 703;  
Best Local Similarity 88.2%; Pred. No. 2e+03; Indels 0; Gaps 0;  
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACCAGTSCCGGCATG 18  
|||||:|||||  
Db 496 ACCAGTCCCGGCATG 512

RESULT 4  
CN225729 738 bp mRNA linear EST 09-APR-2004  
LOCUS WIA077C07.abi Wtbrain Gallus gallus cDNA 5', mRNA sequence.  
DEFINITION CN225729  
ACCESSION CN225729  
VERSION CN225729.1 GI:46329220  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 738)  
Savolainen, P., Fitzsimmons, C.J., Arvestad, L., Andersson, L. and  
Lundeberg, J.  
EST analysis of brain and testis cDNA libraries from White Leghorn  
and Red Jungle Fowl  
Unpublished (2004)  
Contact: Peter Savolainen  
Department of Biotechnology  
Royal Institute of Technology, KTH  
SE-106 91 Stockholm, SWEDEN  
Tel: +46 (0)8 5537 8481  
Fax: +46 (0)8 5537 8335  
Email: [Peter.Savolainen@biotech.kth.se](mailto:Peter.Savolainen@biotech.kth.se)  
Seq primer: M13 reverse primer.  
Location/Qualifiers  
1..738  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="White Leghorn"  
/db\_xref="taxon:9031"  
/sex="female"  
/lab\_host="Wtbrain"  
/clone\_lib="Wtbrain"  
/note="Organ: brain; Vector: pSPORT-1; Site\_1: Hind III;  
Site\_2: EcoRI; The cDNA libraries were created with the  
Superscript Plasmid System (Invitrogen)."

ORIGIN  
Query Match 90.0%; Score 16.2; DB 7; Length 738;  
Best Local Similarity 86.2%; Pred. No. 2e+03; Indels 0; Gaps 0;  
Matches 15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACCAGTSCCGGCATG 18  
|||||:|||||  
Db 418 ACCAGTCCCGGCATG 434

RESULT 5  
BF813258 223 bp mRNA linear EST 12-JAN-2001  
LOCUS PMA-C10113-251100-005-d06 C10113 Homo sapiens cDNA, mRNA sequence.  
DEFINITION BF813258  
ACCESSION BF813258  
VERSION BF813258.1 GI:12144019  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 223)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Negai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matukuma, A., Bata, G.S., Simpson, D.H.,  
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V.,  
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.



TITLE	JOURNAL	MEDLINE	PUBMED	COMMENT
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags	Proc. Natl. Acad. Sci. U.S.A.	97 (7), 3491-3496	(2000)	
20202863				
10737800				
Contact: Stimpson A.J.G.				
Laboratory of Cancer Genetics				
Ludwig Institute for Cancer Research				
Rua Prof. Antonio Prudente, 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil				
Tel: +55-11-2704922				
Fax: +55-11-2707001				
Email: asimpson@ludwig.org.br				
This sequence was derived from the FAPESP/RIKEN Human Cancer Genome Project. This entry can be seen in the following URL				
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=PM4&t2=PM4-C10113-251100-005-d06&t3=2000.11-25&t4=1)				
Seq primer: puc 18 forward				
High quality sequence start: 16				
High quality sequence stop: 97.				
Location/Qualifiers				
1..223				
/organism="Homo sapiens"				
/mol_type="mRNA"				
/db_xref="taxon:9606"				
/dev_stage="Adult"				
/clone_id="C10113"				
/note="Organ: colon, ins; Vector: puc18; Site 1: Sma1; Site 2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."				
Query Match	86.74;	Score 15.6;	DB 2;	Length 223;
Best Local Similarity	83.3%;	Pred. NO. 4e+03;		
Matches	15;	Conservative 2;	Mismatches 1;	Indels 0;
				Gaps 0;
1	GACCACGTCGCGGCATG	18		
145	GACCACGTGCGGCATG	128		
RESULT 6				
BB397065				
LOCUS				
DEFINITION	BB397065	293 bp	mRNA	linear
ACCESSION	BB397065			EST 15-JUL-2000
VERSION	BB397065			
KEYWORDS	BB397065.1	GI:9216461		
SOURCE				
ORGANISM	Mus musculus	(house mouse)		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	Komono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Chiriac, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hironaka, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigenoto, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suganara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamana, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.			
TITLE	RIKEN Mouse ESTs (Komono, H., et al.)			
JOURNAL	Unpublished (2000)			

**COMMENT**

Contact: Yoshinobu Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-resgsc.riken.jp URL:<http://genome.gsc.riken.jp/>  
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagasaki,S., Sasakata,N., Okaaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Thermotubularization and thermotubularization of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoch,M., Katsunai,T., Akiyama,Y., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okaaki,Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 13-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

**FEATURES**

**Source**

Location/Qualifiers

1..293

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="CG30002B21"

/cell\_type="ES cells"

/lab\_host="SOLR"

/note="RIKEN full-length enriched, ES cells"

/cote="Site 1: XhoI; Site 2: BamHI; CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5']  
GAGAGAGAGATCTCGAGTAATTAAATATATCCCCCCCCCCC 3'"]."

**ORIGIN**

Query Match 86.7%; Score 15.6; DB 2; Length 293;  
Best Local Similarity 83.3%; Prod. No. 4e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGCCGGCATG 18  
|||||:::|||||  
Db 44 GACCACGTCCCGGCATG 61

**RESULT 7**

CB995647 295 bp mRNA linear EST 01-MAY-2003  
LOCUS ABENCOUR\_13617193 NIH MGC 148 Homo sapiens cDNA clone IMAGE:3039972 5', mRNA sequence.

**ACCESSION**

CB995647  
VERSION CB995647.1 GI:30290167

**KEYWORDS**

EST.

**SOURCE**

Homo sapiens (human)

**ORGANISM**

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

**REFERENCE**

1 (bases 1 to 295)  
NIH-MGC <http://mgc.ncl.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)

**TITLE**

JOURNAL COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-rtmail.nih.gov

Tissue Procurement: Dr. Stefan Hansson  
cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help and advice from Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULM)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ULM at:  
<http://image.llnl.gov>  
Plate: NDM369 row: b column: 13  
High quality sequence stop: 295.  
Location/Qualifiers

FEATURES  
source 1..295

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3039972"  
/tissue\_type="pre-ecclampic placenta"  
/lab\_host="DH10B TONa"  
/clone\_lib="NIH MGC 148"  
/note="Organ: Placenta; Vector: pBluescriptR; Site 1: all-xhoI; Site 2: BamH; Library is oligo-dT primed and directionally cloned using primer 5'-TTTTTCTTTTCTTCTT-3', size-selected for average insert size 2.3 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH MGC library."

## ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 295;  
Best Local Similarity 83.3%; Pred. No. 4e+03; 1; Indels 0; Gaps 0;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGSCCGGCATG 18  
|||||:|:|:|:|:|  
171 GACCACTGSCCGGCATG 188

RESULT 8 433 bp mRNA linear EST 29-DEC-2000  
BF704768

LOCUS F01.51.C08.b1.A003 Floral-induced Meristem 1 (F01) Sorghum  
DEFINITION F01.51.C08.b1.A003 Floral-induced Meristem 1 (F01) Sorghum  
proingnum cDNA, mRNA sequence.

ACCESSION BF704768  
VERSION BF704768.1 GI:11996429

KEYWORDS EST.

SOURCE Sorghum propinquum

ORGANISM Sorghum propinquum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 433)  
Cordonnier-Pratt, M.-M., Gingle, A., Sudman, M., Marsala, C. and Pratt, L.H.

TITLE An EST database from Sorghum: floral-induced meristems  
JOURNAL Unpublished (2000)  
COMMENT Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.  
Seq primer: JEN REV  
High quality sequence stop: 429

FEATURES  
source POLYA=No. Location/Qualifiers

1..433  
/organism="Sorghum propinquum"  
/mol\_type="mRNA"  
/db\_xref="taxon:132711"  
/clone\_lib="Floral-Induced Meristem 1 (F01)"  
/note="Organ: Floral-Induced meristems; Vector: pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; mature plants were placed in a growth chamber for 15 days with 16 hr darkness and 8 hr light (flowering is induced by short-day conditions); 16 days after being returned to the greenhouse under natural long days during late April/early May, meristems were harvested The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

## ORIGIN

Query Match 86.7%; Score 15.6; DB 2; Length 433;  
Best Local Similarity 83.3%; Pred. No. 4e+03; 1; Indels 0; Gaps 0;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGSCCGGCATG 18  
|||||:|:|:|:|:|  
334 GCCCACTGSCCGGCATG 317

RESULT 9 435 bp mRNA linear EST 20-DEC-2000  
BF656537

LOCUS F01.51.C08.g1.A003 Floral-Induced Meristem 1 (F01) Sorghum  
DEFINITION F01.51.C08.g1.A003 Floral-Induced Meristem 1 (F01) Sorghum  
proingnum cDNA, mRNA sequence.

ACCESSION BF656537  
VERSION BF656537.1 GI:11921671

KEYWORDS EST.

SOURCE Sorghum propinquum

ORGANISM Sorghum propinquum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 435)  
Cordonnier-Pratt, M.-M., Gingle, A., Sudman, M., Marsala, C. and Pratt, L.H.

TITLE An EST database from Sorghum: floral-induced meristems  
JOURNAL Unpublished (2000)  
COMMENT Contact: Cordonnier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.  
Seq primer: T7  
High quality sequence start: 22  
High quality sequence stop: 435  
POLYA=Yes.

FEATURES  
source Location/Qualifiers

1..435  
/organism="Sorghum propinquum"  
/mol\_type="mRNA"  
/db\_xref="taxon:132711"  
/clone\_lib="Floral-Induced Meristem 1 (F01)"  
/note="Organ: Floral-Induced meristems; Vector: pBluescript II from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; mature plants were placed in a growth chamber for 15 days with 16 hr darkness and 8 hr light (flowering is induced by short-day conditions); 16 days after being returned to the greenhouse under natural long days during late April/early May, meristems were harvested The

ORIGIN  
library was made from poly-A RNA in the cloning vector  
lambda ZAP II. Clones to be sequenced were prepared by  
mass excision."

Query Match 86.7%; Score 15.6; DB 2; Length 435;  
Best Local Similarity 83.3%; Pred. No. 4e+03; 1; Indels 0; Gaps 0;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 GACCACGTSCCGGCATG 18  
| | | | | | | | | | | | | | | | | | | | | |  
| | | | | | | | | | | | | | | | | | | | | |  
Db 204 GCCCAGCTCCCGGCATG 187

RESULT 10  
BB729006 451 bp mRNA linear EST 12-OCT-2001  
LOCUS BB729006 RIKEN full-length enriched, 8 cells embryo Mus musculus  
DEFINITION CNA clone B660104115 3', mRNA sequence.  
ACCESSION BB729006  
VERSION BB729006.1 GI:16112281  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM

REFERENCE  
AUTHORS  
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Eukaryota; Euteleostomi; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 451)  
Akimura,T., Aikawa,T., Carninci,P., Furuno,M., Hangaki,T.,  
Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K.,  
Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,  
Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,  
Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,  
Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,  
Sugabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaki-Akaiura,S.,  
Tanaka,T., Tomaru,A., Toya,T., Watanabe,A., Yasunishi,A.,  
Muramatsu,M. and Hayashizaki,Y.  
RIKEN Encyclopedia of Mouse Full-length cDNAs (AKimura,T., et al.  
2001)  
TITLE  
JOURNAL  
COMMENT  
Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.jp, URL: http://genome.gsc.riken.jp/  
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,  
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
waghi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,  
Matsuda,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,  
Matsuyama,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.  
and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,  
Sugahara,Y. and Hayashizaki,Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.  
e mouse tissues.  
Location/Qualifiers  
1. 451  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"

## FEATURES

source  
1. 451  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"

ORIGIN  
/clone="B660104115"  
/cell\_type="8 cells embryo"  
/dev\_stage="8 cells embryo"  
/clone\_1b="RIKEN full-length enriched, 8 cells embryo"

Query Match 86.7%; Score 15.6; DB 2; Length 451;  
Best Local Similarity 83.3%; Pred. No. 4e+03; 1; Indels 0; Gaps 0;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 GACCACGTSCCGGCATG 18  
| | | | | | | | | | | | | | | | | | | | | |  
| | | | | | | | | | | | | | | | | | | | | |  
Db 28 GACCCAGCTCCCGGCATG 11

RESULT 11  
CL797916 510 bp DNA linear GSS 06-AUG-2004  
LOCUS CL797916 OR\_CBA0008006.r OR\_CBA Oryza rufipogon genomic clone OR\_CBA0008006  
DEFINITION 3', genomic survey sequence.  
ACCESSION CL797916  
VERSION CL797916.1 GI:51020275  
KEYWORDS GSS.  
SOURCE Oryza rufipogon  
ORGANISM

REFERENCE  
AUTHORS  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; Oryza.  
1 (bases 1 to 510)  
Kim,H., Yu,Y., Wisotski,M., Yost,D., Stum,D., Rao,K., Luo,M.,  
Jettly,R., Kudrna,D., Muller,C., Halfield,J., Soderlund,C. and  
Wing,R.  
OMAP project  
Unpublished (2004)  
Contact: Rod A. Wing  
Arizona Genomics Institute  
University of Arizona  
Forbes Building Room 303, Tucson, AZ 85721-0036, USA  
Tel: 520 626 9595  
Fax: 520 621 1259  
Email: http://genome.arizona.edu  
PCR Primers  
FORWARD: TAA TAC GAC TCA CTA TAG GG  
BACKWARD: CAC TCA TTA GGC ACC CCA  
Plate: 0008 row: J column: 06  
Seq primer: CAC TCA TTA GGC ACC CCA  
Class: BAC ends.  
Location/Qualifiers  
1. 510  
/organism="Oryza rufipogon"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4529"  
/clone="OR\_CBA0008006"  
/tissue\_type="young leaves"  
/dev\_stage="2 week old seedlings"  
/lab\_host="DH10B T1 phage resistant"  
/clone\_1b="OR\_CBA"  
/note="Vector: pGIBAC1, Site\_1: HindIII, Site\_2: HindIII;  
dir treated 36 hrs before harvest"

## FEATURES

source  
1. 510  
/organism="Oryza rufipogon"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4529"  
/clone="OR\_CBA0008006"  
/tissue\_type="young leaves"  
/dev\_stage="2 week old seedlings"  
/lab\_host="DH10B T1 phage resistant"  
/clone\_1b="OR\_CBA"  
/note="Vector: pGIBAC1, Site\_1: HindIII, Site\_2: HindIII;  
dir treated 36 hrs before harvest"

## ORIGIN

Query Match 86.7%; Score 15.6; DB 9; Length 510;  
Best Local Similarity 83.3%; Pred. No. 4e+03; 1; Indels 0; Gaps 0;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 GACCACGTSCCGGCATG 18  
| | | | | | | | | | | | | | | | | | | | | |  
| | | | | | | | | | | | | | | | | | | | | |  
Db 335 GACCACGTGCGCGCATG 352

RESULT 12  
AQ397255/c 514 bp DNA linear GSS 06-MAR-1999  
LOCUS AQ397255

```

COMMENT          JOURNAL
                  TITLE
                  Exelixis FlyTag EST Project CK02 Library
                  Unpublished (2004)
                  Contact: Stapleton, M.
                  BDGP
                  Lawrence Berkeley National Lab
                  One Cyclotron Rd, Berkeley, CA 94720, USA
                  Fax: 510 486 6798
                  Email: http://www.fruitfly.org/EST\_est@fruitfly.berkeley.edu
                  Plate: BP 185 row: A column: 12
                  High quality sequence stop: 401.
                  Location/Qualifiers
                  1..523
FEATURES
  source
    /organism="Drosophila melanogaster"
    /mol_type="mRNA"
    /db_xref="taxon:7227"
    /clone="BP18512"
    /clone_1b="Exelixis FlyTag CK02 pCDNA-SK+"
    /note="Organ: mixed stage embryo/EST, imaginal disks, and
    adult heads; Vector: pCDNA-SK+; Site 1: NotI; Site 2:
    XhoI; Random primed, normalized library from mixed stage
    embryos, imaginal disks, and adult heads. Subset of
    Exelixis FlyTag CK01 clones sequenced from 3' end"
ORIGIN
Query Match          86.7%; Score 15.6; DB 7; Length 523;
Best Local Similarity 83.3%; Pred. No. 4e+03;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY      1 GACCACGTCSCCGGCATG 18
        |||||:|:|:|:|:|:|
Db       255 GTCACGTCGCCGCGCATG 238

RESULT 14
LOCUS      AL903319          539 bp      mRNA      linear      EST 02-JUN-2004
DEFINITION AL903319 PIR-Z1+Z2 Danio rerio cDNA clone 001-B06-1, mRNA sequence..
ACCESSION  AL903319
VERSION     AL903319.1  GI:23165878
KEYWORDS    EST.
SOURCE      Danio rerio (zebrafish)
ORGANISM    Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 539)
            Lo,J., Lee,S., Xu,M., Liu,F., Ruan,H., Eun,A., He,Y., Ma,W.,
            Wang,W., Wen,Z. and Peng,J.
            15000 unique zebrafish EST clusters and their future use in
            microarray for profiling gene expression patterns during
            embryogenesis
            EmbryoGenesis
            Genome Res. 13 (3), 455-466 (2003)
JOURNAL     22505427
MEDLINE     12618376
PUBMED
COMMENT     Contact: Peng J
            Lab of Functional Genomics
            Institute of Molecular and Cell Biology
            30 Medical Drive, Singapore, 117609, Singapore
            Email: pengji@imcb.a-star.edu.sg
            Clone requests: info@embiosystems.com
            Open Biosystems,
            6705 Odysseey Drive, Huntsville, AL 35806.
            Location/Qualifiers
            1..539
FEATURES
  source
    /organism="Danio rerio"
    /mol_type="mRNA"
    /strain="local wildtype"
    /db_xref="taxon:7955"
    /clone="001-B06-1"
    /tissue_type="whole embryo or fish"
    /dev_stage="mixed stages"
    /clone_1id="PIR-Z1+Z2"
ORIGIN

```

Query Match 86.7%; Score 15.6; DB 1; Length 539;  
 Best Local Similarity 83.3%; Pred. No. 4e+03;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GACCACGTCGCCGCATG 18  
 |||||:|||||  
 Db 260 GACCACGTCGCCGCATG 277

RESULT 15  
 LBAFB3C09/C 564 bp DNA linear GSS 20-JUN-2003  
 LOCUS Leishmania braziliensis GSS, clone LBAFB3C09, genomic survey  
 DEFINITION  
 Sequence.  
 BX544274  
 VERSION BX544274.1 GI:32139513  
 KEYWORDS GSS; genomic survey sequence.  
 SOURCE Leishmania braziliensis  
 ORGANISM Leishmania braziliensis  
 Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;  
 Leishmania; Leishmania braziliensis species complex.

REFERENCE 1  
 LAURENTINO, E.C., RUIZ, J.C. and CRUZ, A.K.  
 GSS analysis of the Leishmania braziliensis genome  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 564)  
 AUTHORS Cruz, A.K.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-JUN-2003) Cruz A.K., University of Sao Paulo,  
 Department of Molecular and Cell Biology, FMRP, Avenida  
 Bandeirantes 3900, Ribeirao Preto, SP, 14049-900, BRAZIL  
 COMMENT Clone requests: akcruz@fmrp.usp.br.  
 FEATURES  
 source  
 1. 564  
 /organism="Leishmania braziliensis"  
 /mol\_type="genomic DNA"  
 /strain="MHOM/BR/75/M2904"  
 /db\_xref="taxon:5660"  
 /clone="LBAFB3C09"

ORIGIN  
 Query Match 86.7%; Score 15.6; DB 9; Length 564;  
 Best Local Similarity 83.3%; Pred. No. 4e+03;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GACCACGTCGCCGCATG 18  
 |||||:|||||  
 Db 355 GACCACGTCGCCGCATG 338

RESULT 16  
 BP221909 581 bp mRNA linear EST 15-SEP-2004  
 LOCUS BP221909 Sugano cDNA library, COL0679 Homo sapiens cDNA clone  
 DEFINITION  
 COR00491, mRNA sequence.  
 ACCESSION BP221909  
 VERSION BP221909.1 GI:52094814  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 581)  
 Suzuki, Y., Yamashita, R., Shirota, M., Sakakibara, Y., Chiba, J.,  
 Mizushima-Sugano, J., Nakai, K. and Sugano, S.  
 Sequence comparison of human and mouse genes reveals a homologous  
 block structure in the promoter regions  
 Genome Res. 14 (9), 1711-1718 (2004)  
 CONTACT: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

FEATURES  
 source  
 Location/Qualifiers  
 1. 581  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="COR00491"  
 /cell\_line="COL0679"  
 /clone\_1lb="Sugano cDNA library, COL0679"  
 /note="melanoma"

ORIGIN  
 Query Match 86.7%; Score 15.6; DB 5; Length 581;  
 Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GACCACGTCGCCGCATG 18  
 |||||:|||||  
 Db 117 GACCACGTCGCCGCATG 134

RESULT 17  
 BP356698 582 bp mRNA linear EST 17-SEP-2004  
 LOCUS BP356698 Sugano cDNA library, mammary gland T47D Homo sapiens cDNA  
 DEFINITION  
 clone TDR01021, mRNA sequence.  
 ACCESSION BP356698  
 VERSION BP356698.1 GI:52286705  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 582)  
 Suzuki, Y., Yamashita, R., Shirota, M., Sakakibara, Y., Chiba, J.,  
 Mizushima-Sugano, J., Nakai, K. and Sugano, S.  
 Sequence comparison of human and mouse genes reveals a homologous  
 block structure in the promoter regions  
 Genome Res. 14 (9), 1711-1718 (2004)  
 CONTACT: Yutaka Suzuki  
 Department of Virology  
 Institute of Medical Science, University of Tokyo  
 4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
 Email: yusuzuki@ims.u-tokyo.ac.jp.

FEATURES  
 source  
 1. 582  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="TDR01021"  
 /tissue\_type="mammary gland"  
 /cell\_line="T47D"  
 /clone\_1lb="Sugano cDNA library, mammary gland T47D"  
 /note="mammary gland tumor"

ORIGIN  
 Query Match 86.7%; Score 15.6; DB 5; Length 582;  
 Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 GACCACGTCGCCGCATG 18  
 |||||:|||||  
 Db 163 GACCACGTCGCCGCATG 180

RESULT 18  
 BP364560 582 bp mRNA linear EST 17-SEP-2004  
 LOCUS BP364560 Sugano cDNA library, fetal lung fibroblast TIG Homo  
 DEFINITION  
 sapiens cDNA clone TIR04501, mRNA sequence.  
 ACCESSION BP364560  
 VERSION BP364560.1 GI:52294802  
 KEYWORDS EST.

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,  
Mizushima-Sugano,J., Nakai,K. and Sugano,S.  
Sequence comparison of human and mouse genes reveals a homologous  
block structure in the promoter regions  
JOURNAL Genome Res. 14 (9), 1711-1718 (2004)  
COMMENT Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yusuzuki@ims.u-tokyo.ac.jp.  
Location/Qualifiers

FEATURES  
source  
1..582  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="TIR04501"  
/issue\_type="lung"  
/cell\_type="fibroblast"  
/dev\_stage="fetal"  
/clone\_lib="Sugano cDNA library, fetal lung fibroblast  
TIG"

ORIGIN  
Query Match 86.7%; Score 15.6; DB 5; Length 582;  
Best Local Similarity 83.3%; Pred.No.4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
1 GACCACGTCSCCGGCATG 18  
|||||:|||||  
144 GACCACGGCCCGGCATG 161

DB  
BP280394 583 bp mRNA linear EST 16-SEP-2004  
BP280394 Sugano cDNA library, KG-1-C Homo sapiens cDNA clone  
BP280394 mRNA sequence.  
BP280394  
BP280394.1 GI:52194126  
EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,  
Mizushima-Sugano,J., Nakai,K. and Sugano,S.  
Sequence comparison of human and mouse genes reveals a homologous  
block structure in the promoter regions  
JOURNAL Genome Res. 14 (9), 1711-1718 (2004)  
COMMENT Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yusuzuki@ims.u-tokyo.ac.jp.  
Location/Qualifiers

FEATURES  
source  
1..583  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="KGR07789"  
/cell\_line="KG-1-C"  
/clone\_lib="Sugano cDNA library, KG-1-C"  
/note="glioma"

ORIGIN  
Query Match 86.7%; Score 15.6; DB 5; Length 583;  
Best Local Similarity 83.3%; Pred.No.4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
1 GACCACGTCSCCGGCATG 18  
|||||:|||||  
144 GACCACGGCCCGGCATG 161

Best Local Similarity 83.3%; Pred.No.4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
1 GACCACGTCSCCGGCATG 18  
|||||:|||||  
136 GACCACGGCCCGGCATG 153

DB  
BP359091 583 bp mRNA linear EST 17-SEP-2004  
BP359091 Sugano cDNA library, mammary gland T47D Homo sapiens cDNA  
BP359091 mRNA sequence.  
BP359091  
BP359091.1 GI:52289104  
EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,  
Mizushima-Sugano,J., Nakai,K. and Sugano,S.  
Sequence comparison of human and mouse genes reveals a homologous  
block structure in the promoter regions  
JOURNAL Genome Res. 14 (9), 1711-1718 (2004)  
COMMENT Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo  
4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yusuzuki@ims.u-tokyo.ac.jp.  
Location/Qualifiers

FEATURES  
source  
1..583  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="TDR07864"  
/issue\_type="mammary gland"  
/cell\_line="T47D"  
/clone\_lib="Sugano cDNA library, mammary gland T47D"  
/note="mammary gland tumor"

ORIGIN  
Query Match 86.7%; Score 15.6; DB 5; Length 583;  
Best Local Similarity 83.3%; Pred.No.4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
1 GACCACGTCSCCGGCATG 18  
|||||:|||||  
136 GACCACGGCCCGGCATG 153

DB  
BP364824 587 bp mRNA linear EST 17-SEP-2004  
BP364824 Sugano cDNA library, fetal lung fibroblast TIG Homo  
sapiens cDNA clone TIR05159, mRNA sequence.  
BP364824  
BP364824.1 GI:52295066  
EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE Suzuki,Y., Yamashita,R., Shiota,M., Sakakibara,Y., Chiba,J.,  
Mizushima-Sugano,J., Nakai,K. and Sugano,S.  
Sequence comparison of human and mouse genes reveals a homologous  
block structure in the promoter regions  
JOURNAL Genome Res. 14 (9), 1711-1718 (2004)  
COMMENT Contact: Yutaka Suzuki  
Department of Virology  
Institute of Medical Science, University of Tokyo

4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan  
Email: yusuzuki@ims.u-tokyo.ac.jp.

# FEATURES

## SOURCE

1. 587

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="TIR05159"

/tissue\_type="lung"

/cell\_type="fibroblast"

/cell\_line="f1g"

/dev\_stage="fetal"

/clone\_lib="Sugano CDNA library, fetal lung fibroblast

TIG"

# VERSION

AZ411523.1 GI:10535536

# KEYWORDS

GSS.

# SOURCE

Mus musculus (house mouse)

# ORGANISM

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

# REFERENCE

1 (bases 1 to 588)

# AUTHORS

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Irlam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.

# TITLE

Muscle whole genome scaffolding with paired end reads from 10kb

# JOURNAL

Unpublished (2000)

# COMMENT

Contact: Robert B. Weiss  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: rdunn@genetics.utah.edu  
Insert length: 10000 Std Error: 0.00  
Plate: 0184 row: C column: 11  
Seq primer: CACACAGGAAACAGCTATGACC  
Class: plasmid ends  
High quality sequence stop: 588.

# FEATURES

Location/Qualifiers

# SOURCE

1. 588

# ORGANISM

/organism="Mus musculus"

# MOL

/mol\_type="genomic DNA"

# DB

/db\_xref="taxon:10090"

# CD

/clone="UGGCM0184C11"

# SEX

/sex="Male"

# LIB

/lib\_host="E. Coli strain XL10-Gold, T1-resistant, F-"

# CLONE

/clone\_lib="Mouse 10kb plasmid UGGCM library"

# NOTE

/note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource

# URL

(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The

# ADAPTOR

adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g14732114[gb|AF129072.1], a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptor complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into

# CHEM

chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

# ORIGIN

Query Match 86.7%; Score 15.6; DB 8; Length 588;

Best Local Similarity 83.3%; Pred. No. 4.1e+03;

Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

1 GACCAGTSCCGGCGCATG 18

|||||:|||||

Db 575 GACCAGTSCCGGCGCATG 558

# RESULT 24

A1389668 602 bp mRNA linear EST 02-DEC-2003

# LOCUS

GH20904.5 Sprime GH Drosophila melanogaster head P072 Drosophila

# DEFINITION

melanogaster cDNA clone GH20904.5 similar to CG6305; Fpan0006305

# GO

[structural protein (GO:0005198)] located on: 2R 50C9-50C11;

RESULT 23  
AZ411523/c 588 bp DNA linear GSS 03-OCT-2000  
LOCUS  
DEFINITION  
clone UGCC1M0184C11 R, genomic survey sequence.  
ACCESSION  
AZ411523

Query Match 86.7%; Score 15.6; DB 5; Length 588;  
Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

1 GACCAGTSCCGGCGCATG 18

|||||:|||||

Db 136 GACCAGGCGCGGCGCATG 153

# ORIGIN

Query Match 86.7%; Score 15.6; DB 5; Length 588;

Best Local Similarity 83.3%; Pred. No. 4.1e+03;

Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

1 GACCAGTSCCGGCGCATG 18

|||||:|||||

Db 136 GACCAGGCGCGGCGCATG 153

RESULT 23  
AZ411523/c 588 bp DNA linear GSS 03-OCT-2000  
LOCUS  
DEFINITION  
clone UGCC1M0184C11 R, genomic survey sequence.  
ACCESSION  
AZ411523



ACCESSION	08/13/2002, mRNA sequence.			
VERSION	A1389668			
KEYWORDS	A1389668.1 GI:4203679			
SOURCE	EST.			
ORGANISM	Drosophila melanogaster (fruit fly)			
	Drosophila melanogaster			
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
	Ephydroidea; Drosophilidae; Drosophila.			
REFERENCE	1 (bases 1 to 602)			
AUTHORS	Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G., Lewis,S. and Rubin,G.M.			
TITLE	BDGP/HMMI Drosophila EST Project			
JOURNAL	Unpublished (2001)			
COMMENT	Other_ESTs: GH20904.3prime Contact: Stapleton, M. BDGP			
FEATURES	Lawrence Berkeley National Lab One Cyclotron Rd, Berkeley, CA 94720, USA Fax: 510 486 6798 Email: <a href="http://www.fruitfly.org/EST/est@fruitfly.berkeley.edu">http://www.fruitfly.org/EST/est@fruitfly.berkeley.edu</a> hit genome: AB003817; arm:2R [8578716,8640572] estimated-cyto:50B9-50C14: 04/10/2001 Plate: GH.209 row: A column: 4 High quality sequence stop: 509 POLYA=No			
Source	Location/Qualifiers			
	1..602			
	/organism="Drosophila melanogaster"			
	/mol_type="mRNA"			
	/db_xref="taxon:7227"			
	/clone="GH20904"			
	/sex="male and female"			
	/dev_stage="adult"			
	/lab_host="DH5 - alpha"			
	/clone.lib="GH Drosophila melanogaster head pot2"			
	/note="Organ: head; Vector: pot2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into pot2. Plasmid cDNA library."			
ORIGIN				
Query Match	86.7%; Score 15.6; DB 1; Length 602;			
Best Local Similarity	83.3%; Pred. No. 4,le+03;			
Matches	15; Conservative	2; Mismatches	1; Indels	0; Gaps 0;
QY	1 GACCACTGTCGGGCATG 18      : : : :			
Db	454 GTCACGTCGCCGGCATG 471			
RESULT 25				
CD311022/c	612 bp mRNA linear EST 16-SEP-2003			
LOCUS	StrP6u61.004386 Sea urchin larva cDNA library MPM6p691			
DEFINITION	Strongylocentrotus purpuratus cDNA clone			
	MPM6p691H0297;MPI_STRUD1_97H2 5', mRNA sequence.			
ACCESSION	CD311022			
VERSION	CD311022.1 GI:34756071			
KEYWORDS	EST.			
SOURCE	Strongylocentrotus purpuratus			
ORGANISM	Strongylocentrotus purpuratus			
	Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;			
	Echinoida; Euechinoidea; Echinacea; Echinoida;			
	Strongylocentrotidae; Strongylocentrotus.			
REFERENCE	1 (bases 1 to 612)			
AUTHORS	Poustka,A.J., Groth,D., Hennig,S., Thamm,S., Cameron,A., Beck,A., Reinhardt,R., Herwig,R., Panopoulou,G. and Lehrach,H.			
TITLE	Generation, annotation, evolutionary analysis, and database			
JOURNAL	Genome Res. 13 (12), 2736-2746 (2003)			
COMMENT	Contact: Poustka AJ laboraty 145, dept.lehrach Max-Planck-Institut fuer Molekulare Genetik			

The library was characterised by oligonucleotide fingerprinting (ONF) to reduce sequencing redundancy. According to the ONF procedure, clones that display the same hybridisation matrix with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per ONF cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well as the coordinates of the other clones assigned to the same ONF cluster as the clone from which the above EST is generated is available at the sea urchin project web site at: [http://www.molgen.mpg.de/ag\\_seaurchin/](http://www.molgen.mpg.de/ag_seaurchin/). cDNA clones and filters are distributed via the Resource Center/Primary Database of the German Human Genome Project (<http://www.rzpd.de>)  
 PCR Primers  
 FORWARD: 5' CCCGAGCTTTACACTTATGCTTCGCGCTCG 3' (M13RSP) 5'-seq  
 BACKWARD: 5' GCTATTACCGCAGCTGCGAAGAGGGAGATGTG 3' (M13FSP) 3'-seq  
 Seq primer: 5' -CCGGTCGCGAATTCGCGGT-3' pSPORT3/86  
 High quality sequence stop: 612.  
 Location/Qualifiers  
 1. 612  
     /organism="Strongylocentrotus purpuratus"  
     /mol\_type="mRNA"  
     /db\_xref="taxon:7668"  
     /cldone="MPMGp691H0297/MP1\_SURUDI\_97H2"  
     /clisue\_type="whole larva"  
     /dev\_stage="larva 2-3 weeks"  
     /lab\_host="E.coli, XL1 blue"  
     /notes="Vector: pSPORT; Site 1: NotI; Site 2: SalI; Random primed and directionally cloned in pSPORT vector using a NotI (5'-GACTGATCTTCAAGTCGCGACGCGCCG (T15-3') and a SalI 5'-TGACCCACGCGCTCG-3' adapters (Gibco BRL)"



Email: cdna@igsun.grc.nia.nih.gov  
 Plate: E0808 row: A Column: 04  
 Seq primer: M13 Reverse  
 High quality sequence stop: 624  
 POLYA=NO.

# FEATURES

source

1. 624  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:E0808A04-5"  
 /db\_xref="taxon:10090"  
 /clone="NIA:E0808A04 IMAGE:30906339"  
 /issue\_type="4-cell stage embryo"  
 /dev\_stage="4-cell"  
 /lab\_host="DH10B"  
 /clone\_1b="NIA Mouse four-cell-Embryo cDNA library (long)"  
 /note="Vector: pCMV-SPORT6 (Invitrogen); Site\_1: SalI; Site\_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). PMID: 11544199). The mRNAs were extracted from a pool of 360 embryos at 4-cell stage. Double-stranded cDNAs were synthesized with an Oligo(dt) primer (Invitrogen): 5'-TGACTAGTTCATGATCCGAGCGCCGCTTTTCTTTT-3' from 10.8ng of mRNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L1-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pCMV-SPORT6 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2Kb. The library was constructed by Yulan Piao."

## ORIGIN

Query Match 86.7%; Score 15.6; DB 7; Length 624;  
 Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGSCCGGCATG 18  
 |||||:|||||  
 Db 100 GACCACTGSCCGGCATG 117

## RESULT 27

CA100015

LOCUS

CA100015 631 bp mRNA linear EST 23-SEP-2003  
 SCVPC16044C07.G Cl6 Saccharum officinarum cDNA clone SCVPC16044C07  
 5', mRNA sequence.

ACCESSION

CA100015

VERSION

CA100015.1

GI:34953322

KEYWORDS

EST.

SOURCE

ORGANISM

Saccharum officinarum

Saccharum officinarum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum

complex.

1 (bases 1 to 631)

Vector, A.L., da Silva, F.R., Kemper, E.L. and Arruda, P.

The libraries that made SUCEST

Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

Contact: Arruda P

Centro de Biologia Molecular e Engenharia Genetica

Universidade Estadual de Campinas

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

# FEATURES

source

Caixa Postal 6010, 13083-970, Campinas SP, Brazil  
 Tel: 55 19 3788 1137  
 Fax: 55 19 3788 1089  
 Email: parryda@unicamp.br  
 Clone distribution information can be found  
 through the Brazilian Clone Collection Center (BCCC) at  
 http://www.bcccenter.fcav.unesp.br  
 Plate: 044 row: C column: 07  
 Seq primer: T7 Promoter Primer.  
 Location/Qualifiers  
 1. 631  
 /organism="Saccharum officinarum"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4547"  
 /clone="SCVPC16044C07"  
 /lab\_host="DH10B"  
 /clone\_1b="Cl6"  
 /note="Organ: Pool of sugarcane calli submitted to low (40C) and high (37 C) temperature stress; Vector: pSPORT1; Site\_1: SalI; Site\_2: NotI; An unidirectional cDNA library generated from (pool of sugarcane calli submitted to low (40C) and high (37 C) temperature stress). cDNA was prepared from polyA+ mRNA using SuperScript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-column and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at http://sucest.lad.ic.unicamp.br/public"

## ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 631;  
 Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGSCCGGCATG 18  
 |||||:|||||  
 Db 300 GACCACTGSCCGGCATG 317

## RESULT 28

CA071958

LOCUS

CA071958 634 bp mRNA linear EST 23-SEP-2003  
 SCBGAM1092E04.G AM1 Saccharum officinarum cDNA clone SCBGAM1092E04  
 5', mRNA sequence.

ACCESSION

CA071958

VERSION

CA071958.1

GI:34924097

KEYWORDS

EST.

SOURCE

ORGANISM

Saccharum officinarum

Saccharum officinarum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum

complex.

1 (bases 1 to 634)

Vector, A.L., da Silva, F.R., Kemper, E.L. and Arruda, P.

The libraries that made SUCEST

Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

Contact: Arruda P

Centro de Biologia Molecular e Engenharia Genetica

Universidade Estadual de Campinas

Caixa Postal 6010, 13083-970, Campinas SP, Brazil

Tel: 55 19 3788 1137

Fax: 55 19 3788 1089

Email: parryda@unicamp.br

Clone distribution information can be found

through the Brazilian Clone Collection Center (BCCC) at

http://www.bcccenter.fcav.unesp.br

Plate: 092 row: E column: 04

Seq primer: T7 Promoter Primer.

Location/Qualifiers

1. 634

/organism="Saccharum officinarum"

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

/mol\_type="mRNA"  
/db\_xref="taxon:4547"  
/clone="SCBAM1092E04"  
/lab\_host="DH10B"  
/clone\_lib="AM1"  
/note="Organ: Apical meristem and tissues surrounding of mature plants; Vector: pSPort1; Site\_1: SalI; Site\_2: NotI; An unidirectional cDNA library generated from [Apical meristem and tissues surrounding of mature plants]. cDNA was prepared from polyA+ mRNA using Superscript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a separose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at <http://eucast.lad.ic.unicaamp.br/public>"

## ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 634;  
Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCAAGTSCCGGCATG 18  
|||||:|||||  
Db 168 GACCAAGCGCCGGCATG 185

RESULT 29  
CF367459 636 bp mRNA linear EST 25-AUG-2003  
LOCUS 842208 MARC 3P1G Sus scrofa cDNA 5', mRNA sequence.  
ACCESSION CF367459  
VERSION CF367459.1 GI:34172024  
KEYWORDS EST.  
SOURCE Sus scrofa (pig)  
ORGANISM Sus scrofa

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
1 (bases 1 to 636)  
Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Fox,J.,  
Wise,T.A., Nonnean,D.J., Wray,J.E. and Keele,J.W.  
A second set of porcine ESTs from a pooled-tissue normalized  
library

## JOURNAL

Unpublished (2003)

## COMMENT

Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: [smith@email.marc.usda.gov](mailto:smith@email.marc.usda.gov)  
Single pass sequencing. Bases called with phred v0.020425.c and  
trimmed with the aid of the trim\_alt option. Vector identified with  
cross match v0.990329.  
Plate: SRG8026 row: F column: 4  
Seq primer: GTATACGACTCACTATAGCG.  
Location/Qualifiers

## FEATURES

source

1. 636  
/organism="Sus scrofa"  
/mol\_type="mRNA"  
/db\_xref="taxon:9823"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/clone\_lib="MARC 3P1G"  
/note="Vector: pCDNA3.1; Site 1: EcoRI; Site 2: NotI;  
library made with RNA pooled from multiple tissues  
including brain, liver, muscle, placenta/endometrium,  
ovary, testes, and bone marrow."

## ORIGIN

Query Match 86.7%; Score 15.6; DB 7; Length 636;  
Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCAAGTSCCGGCATG 18  
|||||:|||||  
Db 419 GACCAAGCGCCGGCATG 402

RESULT 30  
CB096877 641 bp mRNA linear EST 27-JAN-2003  
LOCUS AF53-RP\_09 J17 T7 069.ab1 IRRI Drought Stress Panicle Library  
DEFINITION Oryza sativa (indica cultivar-group) cDNA clone C0003305 5' similar  
to Transcription factor BTF3 (RNA polymerase B transcription factor  
3), mRNA sequence.

ACCESSION CB096877  
VERSION CB096877.1 GI:27921065  
KEYWORDS EST.  
SOURCE Oryza sativa (indica cultivar-group)  
ORGANISM Oryza sativa (indica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehretaceae; Oryzaeae; Oryza.  
1 (bases 1 to 641)  
Bennett,J., Arumugam,K., Laiitte,R., Wen,J., Rudd,S. and  
Bruskiewich,R.M.  
IRRI Drought Stress Panicle cDNA library  
unpublished (2002)  
Contact: Richard Bruskiewich  
Biometrics and Bioinformatics Unit  
International Rice Research Institute  
DAPO 7777, Metro Manila, Philippines  
Tel: +63-2-845-0563  
Fax: +63-2-845-0606  
Email: [r.bruskiewich@cgiar.org](mailto:r.bruskiewich@cgiar.org)  
International Rice Information System (IRIS;  
<http://www.iris.irri.org>): D0203304 Assignment of putative function  
to the sequence by S. Rudd of the Munich Information Center for  
Protein Sequences (<http://mips.gsf.de>)  
Plate: 09 row: J column: 17.  
Location/Qualifiers

1. 641  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="IR64"  
/db\_xref="taxon:39946"  
/clone="C0003305"  
/tissue\_type="Panicles"  
/dev\_stage="Flowering"  
/clone\_lib="IRRI Drought Stress Panicle Library"  
/note="Vector: pBluescript II SK+; Water stress was  
applied by not watering for 4 consecutive days. Panicles  
were collected from control (well watered) and stressed  
plants at 2 days before heading, at heading, 50% flowering  
and 4 days after 50% flowering."

## FEATURES

source

1. 641  
/organism="Oryza sativa (indica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="IR64"  
/db\_xref="taxon:39946"  
/clone="C0003305"  
/tissue\_type="Panicles"  
/dev\_stage="Flowering"  
/clone\_lib="IRRI Drought Stress Panicle Library"  
/note="Vector: pBluescript II SK+; Water stress was  
applied by not watering for 4 consecutive days. Panicles  
were collected from control (well watered) and stressed  
plants at 2 days before heading, at heading, 50% flowering  
and 4 days after 50% flowering."

## ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 641;  
Best Local Similarity 83.3%; Pred. No. 4.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCAAGTSCCGGCATG 18  
|||||:|||||  
Db 499 GACCAAGTCCCGGCATG 482

Search completed: July 20, 2005, 18:15:50  
Job time : 2283.57 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2005, 13:19:40 / Search time 296.143 Seconds  
(without alignments)  
359.811 Million cell updates/sec

Title: US-10-017-471b-1

Perfect score: 18

Sequence: 1 gaccacgtccggcagcagc 18

Scoring table: IDENTITY NUC

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Listing first 300 summaries

N\_Geneseq\_16Dec04:\*  
1: geneeqn1980s:\*  
2: geneeqn1990s:\*  
3: geneeqn2000s:\*  
4: geneeqn2001as:\*  
5: geneeqn2001bs:\*  
6: geneeqn2002as:\*  
7: geneeqn2002bs:\*  
8: geneeqn2003as:\*  
9: geneeqn2003bs:\*  
10: geneeqn2003cs:\*  
11: geneeqn2003ds:\*  
12: geneeqn2004as:\*  
13: geneeqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17.2	95.6	18	6	ABN86366
2	17.2	95.6	18	10	ADBA4795
3	16.2	90.0	18	13	ADQA8866
4	15.6	88.9	4346	6	ABN86379
5	15.6	88.9	4346	10	ADBA4794
6	15.6	86.7	465	3	AACT5505
7	15.6	86.7	513	2	AAAX98247
8	15.6	86.7	612	13	ADBA91782
9	15.6	86.7	744	4	ABLI0881
10	15.6	86.7	766	2	AAAX98245
11	15.6	86.7	891	11	ABDI4835
12	15.6	86.7	1188	10	ABX05949
13	15.6	86.7	1275	11	ABDI5034
14	15.6	86.7	1408	2	AAZ32291
15	15.6	86.7	1408	2	AAZ32287
16	15.6	86.7	2008	8	ADA73004
17	15.6	86.7	2048	2	AAAX98244
18	15.6	86.7	2395	2	AAV52338
19	15.6	86.7	5042	6	ABN59696
20	15.6	86.7	5362	10	ADBO6975

21	15.6	86.7	5624	5	AAH89920	Aah89920 Human bon
22	15.6	86.7	12332	4	ABLI0880	Abli0880 Drosophila
23	15.6	86.7	110000	10	ABBS56454_01	Continuation (2 of
24	15.2	84.4	5356	4	ABLI6420	Abli6420 Drosophila
25	15.2	84.4	110021	13	ABD32859	Abd32859 Mouse can
26	14.6	81.1	209	6	ABLI75000	Abli75000 Corn tass
27	14.6	81.1	520	13	ADOS2850	Adg2850 Novel can
28	14.6	81.1	521	3	AAAC5488	Aac5488 Eucalyptu
29	14.6	81.1	521	3	AAAC56107	Aac56107 Eucalyptu
30	14.6	81.1	568	9	ADA48826	Ada48826 Banana ge
31	14.6	81.1	568	11	ADJ11865	Adj11865 Banana CD
32	14.6	81.1	607	10	ADP33617	Adp33617 Mouse mlt
33	14.6	81.1	888	3	AAZ57117	Aaz57117 Mouse sta
34	14.6	81.1	888	4	AAAF31040	Aaf31040 Murine st
35	14.6	81.1	921	4	ABLI07179	Abli07179 Drosophila
36	14.6	81.1	1005	6	ABN21414	Abn21414 Human ORF
37	14.6	81.1	1041	10	ABZ66670	Abz66670 Orthosomy
38	14.6	81.1	1068	8	ABZ37539	Abz37539 Streptomy
39	14.6	81.1	1092	13	ADT41741	Adt41741 Bacterial
40	14.6	81.1	1209	13	ADBA45831	Adba45831 Bacterial
41	14.6	81.1	1424	4	ABLI02237	Abli02237 Drosophila
42	14.6	81.1	1674	13	ADBS0159	Adbs0159 Bacterial
43	14.6	81.1	1749	13	ADSA46460	Adsa46460 Bacterial
44	14.6	81.1	1749	6	AAAD21928	Aad21928 M. roseni
45	14.6	81.1	2366	11	ADJ118905	Adj118905 Human NOV
46	14.6	81.1	2519	12	ADQ63227	Adq63227 Novel hum
47	14.6	81.1	2534	11	ADM03261	Adm03261 Human CDN
48	14.6	81.1	2996	4	ABLI07178	Abli07178 Drosophila
49	14.6	81.1	3380	4	ABLI07174	Abli07174 Drosophila
50	14.6	81.1	3500	4	ABLI02236	Abli02236 Drosophila
51	14.6	81.1	3799	13	ADSA48391	Adsa48391 Bacterial
52	14.6	81.1	4360	4	ABLI21760	Abli21760 Drosophila
53	14.6	81.1	4979	5	AAAS91509	Aaas91509 DNA encod
54	14.6	81.1	4979	12	ADM32882	Adm32882 Nucleotid
55	14.6	81.1	5874	10	ADL12125	Adl12125 DNA encod
56	14.6	81.1	5874	10	ADL12125	Adl12125 DNA encod
57	14.6	81.1	5997	6	AAI68598	Aai68598 A. cryptu
58	14.6	81.1	6946	8	ABZ24581	Abz24581 Human mal
59	14.6	81.1	7564	10	ADL15056	Adl15056 Human cel
60	14.6	81.1	7616	13	ACNA42378	Acna42378 Human dia
61	14.6	81.1	7838	4	ABLI21408	Abli21408 Drosophila
62	14.6	81.1	8931	3	AAAT75307	Aaat75307 DNA sequ
63	14.6	81.1	11168	6	AAAD21912	Aaad21912 Micromono
64	14.6	81.1	14061	6	ABV93632	Abv93632 Human NOV
65	14.6	81.1	14109	6	ABV93632	Abv93632 Human NOV
66	14.6	81.1	17093	10	ABZ58812	Abz58812 S. cinnam
67	14.6	81.1	20489	12	ADQ22881	Adq22881 Human sof
68	14.6	81.1	24120	8	ABX11642	Abx11642 Human ser
69	14.6	81.1	43865	13	ABD33215	Abd33215 Human can
70	14.6	81.1	45055	10	ABZ66808	Abz66808 Orthosomy
71	14.6	81.1	59816	8	ABZ37516	Abz37516 Streptomy
72	14.6	81.1	59816	8	ABZ37516	Abz37516 Streptomy
73	14.6	81.1	71292	11	ACNA45142	Acna45142 Human gen
74	14.6	81.1	110000	4	AAI99682_34	Continuation (35 o
75	14.6	81.1	110000	4	AAI99682_34	Continuation (35 o
76	14.6	81.1	349980	6	ABO81847	Abog81847 Bifidobac
77	14.6	81.1	349980	6	ABO81848	Abog81848 Bifidobac
78	14.6	80.0	39	6	ABN84675	Abn84675 Vascular
79	14.4	80.0	144	6	ABK3123	Abk3123 Human pro
80	14.4	80.0	654	9	ACH15097	Ach15097 Human adu
81	14.4	80.0	654	9	ABX98613	Abx98613 Rice leaf
82	14.4	80.0	1116	13	ADT41618	Adt41618 Bacterial
83	14.4	80.0	1116	13	ADBS64029	Adbs64029 Bacterial
84	14.4	80.0	1116	13	ADBS64029	Adbs64029 Bacterial
85	14.4	80.0	1185	13	ADT48168	Adt48168 Bacterial
86	14.4	80.0	1644	13	ACNA42143	Acna42143 Human dia
87	14.4	80.0	1981	11	ACNA42142	Acna42142 Human dia
88	14.4	80.0	110000	11	ACNA43998_1	Continuation (2 of
89	14.4	80.0	110000	11	ACNA43998_1	Continuation (2 of
90	14.4	80.0	121062	12	ADQ97313	Adq97313 Human can
91	14.2	78.9	471	6	AAAD35503	Aad35503 Methylomo
92	14.2	78.9	471	6	ABK83262	Abk83262 High grow
93	14.2	78.9	471	6	ABK50085	Methylomo

C 94	14.2	78.9	579	11	ADJ12065	Adj12065 Maize cDN	167	14	77.8	783	11	ACH94524	Ach94524 Klebsiell
C 95	14.2	78.9	593	12	ADQ92250	Adq92250 Human aut	168	14	77.8	789	13	AAI97484	AAI97484 Human neu
C 96	14.2	78.9	796	2	AAV08844	AAV08844 Gene No.	169	14	77.8	789	13	ADSB3110	ADSB3110 Bacterial
C 97	14.2	78.9	897	13	ADS63777	Ads63777 Bacterial	170	14	77.8	789	13	ADSB2648	ADSB2648 Bacterial
C 98	14.2	78.9	897	13	ADS64161	Ads64161 Bacterial	171	14	77.8	789	13	ADSB5467	ADSB5467 Bacterial
C 99	14.2	78.9	909	13	ADS63409	Ads63409 Bacterial	172	14	77.8	840	12	ADNB1501	ADNB1501 Human KPP
C 100	14.2	78.9	1295	3	AAZ40192	Aaz40192 H. vulgar	173	14	77.8	840	13	ADSB5901	ADSB5901 Bacterial
C 101	14.2	78.9	1395	3	AAH78786	AAH78786 Nucleotid	C 174	14	77.8	874	4	AAK93113	AAK93113 Human cDN
C 102	14.2	78.9	1832	6	AAI45918	AAI45918 Terrabact	C 175	14	77.8	874	4	AAK91598	AAK91598 Human cDN
C 103	14.2	78.9	1836	13	ADS64109	Ads64109 Bacterial	C 176	14	77.8	874	12	ADL29540	ADL29540 5' end of
C 104	14.2	78.9	1934	11	ACH95422	Ach95422 Klebsiell	C 177	14	77.8	874	12	ADL28025	ADL28025 5' end of
C 105	14.2	78.9	2035	13	ADS63352	Ads63352 Bacterial	C 178	14	77.8	878	4	AAH03966	AAH03966 Human cDN
C 106	14.2	78.9	2035	13	ADS63352	Ads63352 Bacterial	C 179	14	77.8	885	11	ADRB17960	ADRB17960 Pseudomon
C 107	14.2	78.9	2376	16	ABS69227	Abs69227 Bacterial	C 180	14	77.8	900	13	ADRB1256	ADRB1256 Anti-biof
C 108	14.2	78.9	2627	6	ABS69922	Abs69922 Aspergill	C 181	14	77.8	918	11	ADMA4942	ADMA4942 Insect re
C 109	14.2	78.9	3425	12	ADQ63485	Adq63485 Novel hum	C 182	14	77.8	956	13	ADRB8180	ADRB8180 Aspergill
C 110	14.2	78.9	3702	4	AA541970	AA541970 Genomic s	C 183	14	77.8	979	4	ABA77110	ABA77110 Proteifera
C 111	14.2	78.9	3842	2	AA541971	AA541971 Genomic s	C 184	14	77.8	1041	10	ABZ66675	ABZ66675 Orthosomy
C 112	14.2	78.9	6085	2	AAH70153	AAH70153 S. longisp	C 185	14	77.8	1041	10	ABZ66674	ABZ66674 Orthosomy
C 113	14.2	78.9	9006	5	AAH78783	AAH78783 Nucleotid	C 186	14	77.8	1047	8	ABZ37563	ABZ37563 Streptomy
C 114	14.2	78.9	34719	12	ADP90617	Adp90617 Micromono	C 187	14	77.8	1047	10	ABZ66673	ABZ66673 Orthosomy
C 115	14.2	78.9	110000	11	ADM27081_10	ADM27081_10 o	C 188	14	77.8	1047	10	ABZ66673	ABZ66673 Orthosomy
C 116	14.2	78.8	138	4	ACD93471	ACD93471 Human col	C 189	14	77.8	1089	8	ABZ37517	ABZ37517 Streptomy
C 117	14.2	77.8	139	4	ACD93471	ACD93471 Human col	C 190	14	77.8	1115	3	AAK38229	AAK38229 Arabidops
C 118	14.2	77.8	139	10	ACD27677	ACD27677 Virus/nef	C 191	14	77.8	1124	10	ADK58797	ADK58797 Plant DNA
C 119	14.2	77.8	229	3	AAH41628	AAH41628 Human sec	C 192	14	77.8	1124	13	ADT48637	ADT48637 Bacterial
C 120	14.2	77.8	229	12	ADE52377	Ades52377 Mouse ASP	C 193	14	77.8	1167	2	AAV84507	AAV84507 Human sec
C 121	14.2	77.8	244	8	AAZ4619	AAZ4619 Human gen	C 194	14	77.8	1167	9	ABBA83290	ABBA83290 Human sec
C 122	14.2	77.8	255	8	ABZ37519	Abz37519 Streptomy	C 195	14	77.8	1167	9	ACH04791	ACH04791 Novel hum
C 123	14.2	77.8	268	4	AAI29169	AAI29169 Colon tum	C 196	14	77.8	1167	9	ACD44601	ACD44601 Human cDN
C 124	14.2	77.8	268	8	ABZ33355	Abz333355 Human col	C 197	14	77.8	1167	10	ACF72238	ACF72238 Photorhab
C 125	14.2	77.8	291	2	AAH81111	AAH81111 Nucleotid	C 198	14	77.8	1179	4	AA507129	AA507129 DNA encod
C 126	14.2	77.8	303	8	ACA31973	ACA31973 Prokaryot	C 199	14	77.8	1182	13	ADSA5817	ADSA5817 Bacterial
C 127	14.2	77.8	323	2	AAK97665	AAK97665 Extended	C 200	14	77.8	1218	9	ADA89400	ADA89400 Maize hyp
C 128	14.2	77.8	323	12	ADP18932	ADP18932 Human sec	C 201	14	77.8	1251	11	ABD17850	ABD17850 Pseudomon
C 129	14.2	77.8	335	2	AAK51943	AAK51943 Human sec	C 202	14	77.8	1265	13	ADT47066	ADT47066 Bacterial
C 130	14.2	77.8	339	2	AAK40684	AAK40684 Human sec	C 203	14	77.8	1269	11	ABD17432	ABD17432 Pseudomon
C 131	14.2	77.8	348	4	AAK37034	AAK37034 Novel hum	C 204	14	77.8	1272	6	ABQ90218	ABQ90218 M. capsul
C 132	14.2	77.8	354	8	ABZ71757	Abz71757 Human can	C 205	14	77.8	1288	4	ABL06133	ABL06133 Drosophil
C 133	14.2	77.8	356	6	ABN76664	ABN76664 Human ORF	C 206	14	77.8	1308	8	ACA45189	ACA45189 Prokaryot
C 134	14.2	77.8	411	2	AAK51677	AAK51677 Human sec	C 207	14	77.8	1328	5	AAK89420	AAK89420 DNA encod
C 135	14.2	77.8	413	9	ACH16964	ACH16964 Human adu	C 208	14	77.8	1332	2	AAT90337	AAT90337 Daunorubi
C 136	14.2	77.8	418	9	ACH29501	ACH29501 Human adu	C 209	14	77.8	1356	6	ABK10750	ABK10750 DNA encod
C 137	14.2	77.8	451	8	ABX36177	ABX36177 Bovine ES	C 210	14	77.8	1374	6	ABN70732	ABN70732 Streptoco
C 138	14.2	77.8	464	9	ADB46101	ADB46101 Rice thia	C 211	14	77.8	1382	11	ADMA4532	ADMA4532 Insect re
C 139	14.2	77.8	466	10	ACH43762	ACH43762 Human foe	C 212	14	77.8	1386	11	ADSB56031	ADSB56031 Bacterial
C 140	14.2	77.8	467	9	ACH43762	ACH43762 Human foe	C 213	14	77.8	1389	10	ADG33851	ADG33851 Actinomyc
C 141	14.2	77.8	469	9	ACH45256	ACH45256 Human foe	C 214	14	77.8	1392	10	ADG33845	ADG33845 Actinomyc
C 142	14.2	77.8	477	8	ACH19112	ACH19112 Prokaryot	C 215	14	77.8	1395	10	ADG33849	ADG33849 Actinomyc
C 143	14.2	77.8	490	2	AAK41313	AAK41313 Human sec	C 216	14	77.8	1405	9	ADB46115	ADB46115 Rice thia
C 144	14.2	77.8	495	5	ABA12903	ABA12903 Human ner	C 217	14	77.8	1410	10	ADG33847	ADG33847 Actinomyc
C 145	14.2	77.8	540	10	ADDE6966	ADDE6966 Human lun	C 218	14	77.8	1420	8	ADA72409	ADA72409 Rice gene
C 146	14.2	77.8	540	10	ADDE8220	ADDE8220 Human lun	C 219	14	77.8	1423	10	ADCT75690	ADCT75690 DNA homol
C 147	14.2	77.8	570	11	ABD03664	ABD03664 Pseudomon	C 220	14	77.8	1425	10	ADCT6304	ADCT6304 DNA homol
C 148	14.2	77.8	601	3	AAK35541	AAK35541 Arabidops	C 221	14	77.8	1440	4	AAH52040	AAH52040 Mycobacte
C 149	14.2	77.8	613	5	ABA18510	ABA18510 Human ner	C 222	14	77.8	1440	8	ACA40655	ACA40655 Prokaryot
C 150	14.2	77.8	618	6	ABO66231	ABO66231 Arabidops	C 223	14	77.8	1452	8	ACA38694	ACA38694 Prokaryot
C 151	14.2	77.8	621	12	ADN97658	ADN97658 S. ambofac	C 224	14	77.8	1461	6	ABK10752	ABK10752 DNA encod
C 152	14.2	77.8	636	9	ADB46121	ADB46121 Rice thia	C 225	14	77.8	1476	8	ACA40929	ACA40929 Prokaryot
C 153	14.2	77.8	640	10	ADCT6353	ADCT6353 DNA homol	C 226	14	77.8	1482	8	ACA23810	ACA23810 Prokaryot
C 154	14.2	77.8	644	10	ADCT6353	ADCT6353 DNA homol	C 227	14	77.8	1488	6	ABA92476	ABA92476 Human TIP
C 155	14.2	77.8	648	11	ABD03882	ABD03882 Pseudomon	C 228	14	77.8	1488	10	ADBB44323	ADBB44323 Human rep
C 156	14.2	77.8	660	10	ADCT6356	ADCT6356 DNA homol	C 229	14	77.8	1488	12	ADOB3693	ADOB3693 Human tum
C 157	14.2	77.8	666	6	ABO65868	ABO65868 Arabidops	C 230	14	77.8	1488	13	ADOB4764	ADOB4764 Human tum
C 158	14.2	77.8	671	6	ADN09605	ADN09605 HIV-1 cod	C 231	14	77.8	1488	13	ADOB7002	ADOB7002 Human tum
C 159	14.2	77.8	671	6	ADN71152	ADN71152 Codon opt	C 232	14	77.8	1488	13	ADOB85840	ADOB85840 Human tum
C 160	14.2	77.8	671	10	ACD27660	ACD27660 Humanised	C 233	14	77.8	1488	13	ACN40651	ACN40651 Tumour-as
C 161	14.2	77.8	685	3	AAK51982	AAK51982 Arabidops	C 234	14	77.8	1488	13	ADP54238	ADP54238 Human PRO
C 162	14.2	77.8	728	3	AAK48877	AAK48877 Arabidops	C 235	14	77.8	1492	3	AAA90316	AAA90316 Human HEL
C 163	14.2	77.8	731	3	AAK34504	AAK34504 Arabidops	C 236	14	77.8	1492	4	ABU11047	ABU11047 Drosophil
C 164	14.2	77.8	750	11	ABD06013	ABD06013 Pseudomon	C 237	14	77.8	1494	8	ACA40913	ACA40913 Prokaryot
C 165	14.2	77.8	763	10	ADD26834	ADD26834 Human adi	C 238	14	77.8	1516	10	ADCT3050	ADCT3050 Trameaes
C 166	14.2	77.8	774	10	ACF67152	ACF67152 Photorhab	C 239	14	77.8	1598	10	ADK66965	ADK66965 Gene #55

C 240	14	77.8	1524	6	ABN89320
C 241	14	77.8	1548	4	AAK94131
C 242	14	77.8	1548	12	ADL30597
C 243	14	77.8	1559	8	ACA31456
C 244	14	77.8	1563	4	AAH17993
C 245	14	77.8	1593	8	ACA37741
C 246	14	77.8	1647	11	ABD06454
C 247	14	77.8	1660	6	ABK83216
C 248	14	77.8	1680	6	ABK75035
C 249	14	77.8	1701	11	ADP75211
C 250	14	77.8	1857	8	ACA27195
C 251	14	77.8	1863	8	ABZ72016
C 252	14	77.8	1863	13	ACN37721
C 253	14	77.8	1863	13	ADP54117
C 254	14	77.8	1866	8	ABT32135
C 255	14	77.8	1867	10	ACA92442
C 256	14	77.8	1879	11	AD131416
C 257	14	77.8	1882	11	ADM02350
C 258	14	77.8	1908	2	AAQ99364
C 259	14	77.8	1908	2	AAV84065
C 260	14	77.8	1908	3	AAK61403
C 261	14	77.8	1920	2	AAZ06460
C 262	14	77.8	1920	2	ABL26517
C 263	14	77.8	1920	11	ABD05586
C 264	14	77.8	2023	10	AAV06593
C 265	14	77.8	2023	10	ABSS7455
C 266	14	77.8	2070	11	ABD03585
C 267	14	77.8	2100	8	ADA50514
C 268	14	77.8	2103	8	ACA37871
C 269	14	77.8	2166	12	ADDO0545
C 270	14	77.8	2175	6	ABA99462
C 271	14	77.8	2175	6	ABL90432
C 272	14	77.8	2179	4	AAH17168
C 273	14	77.8	2334	12	ADU40133
C 274	14	77.8	2394	13	ADS48448
C 275	14	77.8	2426	11	ADM02149
C 276	14	77.8	2427	3	AACT7157
C 277	14	77.8	2540	2	AAO51488
C 278	14	77.8	2540	12	ADJ94704
C 279	14	77.8	2582	2	AAT73117
C 280	14	77.8	2672	4	ABL06124
C 281	14	77.8	2712	8	ACA26785
C 282	14	77.8	2817	8	ACA41660
C 283	14	77.8	2827	4	AAAF31253
C 284	14	77.8	2856	13	ADR84593
C 285	14	77.8	2958	5	AA929666
C 286	14	77.8	2960	5	AA90160
C 287	14	77.8	2983	4	AAH14446
C 288	14	77.8	3011	6	ABO99304
C 289	14	77.8	3018	8	AAI61203
C 290	14	77.8	3103	12	ADDO9942
C 291	14	77.8	3103	12	ADQ09868
C 292	14	77.8	3110	12	ADQ24300
C 293	14	77.8	3144	4	AAH47048
C 294	14	77.8	3216	12	AD134850
C 295	14	77.8	3288	4	AB106132
C 296	14	77.8	3561	6	AA97178
C 297	14	77.8	3567	6	ABK12893
C 298	14	77.8	3580	6	AB87817
C 299	14	77.8	3624	9	ACC58643
C 300	14	77.8			

## ALIGNMENTS

RESULT 1  
ID ABN86366 standard; DNA; 18 BP.  
AC ABN86366;  
XX  
XX

DT 21-OCT-2002 (first entry)

XX	S. coelicolor scdA gene internal segment amplifying primer 1.
DE	Antibiotic; bacterium; scdA; afra; scdR; arpa; barA; actinorhodine; Act;
XX	undecylprodigiosin; Red; PCR; primer; ss.
KW	Streptomycetes coelicolor.
OS	CA2322241-A1.
XX	23-APR-2002.
XX	23-OCT-2000; 2000CA-02322241.
PF	23-OCT-2000; 2000CA-02322241.
XX	23-OCT-2000; 2000CA-02322241.
XX	(PLAN-) PLANT BIOSCIENCE LTD.
PA	Takano E, Bibb M;
PI	WPI; 2002-501089/54.
DR	Modifying antibiotic-producing Streptomycetes, to increase, or alter timing
XX	of, antibiotic production, by deleting the scdA or scdR genes.
PT	Claim 19; Page 53; 64pp; English.
XX	The invention provides a method for modifying an antibiotic-producing
CC	strain of Streptomycetes to increase production of antibiotics or to alter
CC	the timing of antibiotic production. The modification is functional
CC	detection of the scdA gene of S. coelicolor, or its homologues, but is not
CC	detection of the afra gene of S. griseus, or the modification is
CC	functional deletion of the scdR gene of S. coelicolor, or its homologues,
CC	but is not deletion of arpa of S. griseus nor barA of S. virginiae. The
CC	method is particularly used for production of the antibiotics
CC	actinorhodine (Act) and undecylprodigiosin (Red). The present sequence
CC	represents a PCR primer for amplifying an internal segment of the scdA
CC	gene from S. coelicolor M145 total DNA
XX	Sequence 18 BP; 3 A; 6 C; 5 G; 2 T; 0 U; 2 Other;
SO	
QY	Query Match 95.6%; Score 17.2; DB 6; Length 18;
DB	Best Local Similarity 100.0%; Pred. No. 1.5e+02;
	Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
	1 GACCACTGCTCCGGCATG 18
	1 GACCACTGCTCCGGCATG 18
RESULT 2	
ID ADE14795 standard; DNA; 18 BP.	
AC ADE14795;	
XX	
XX	29-JAN-2004 (first entry)
DE	PCR primer 1 related to Streptomycetes coelicolor antibiotic production.
XX	antibiotic-producing strain; antibiotic production; scdA gene; scdR gene;
KW	actinorhodin; undecylprodigiosin; PCR; primer; ss.
OS	Synthetic.
XX	Streptomycetes coelicolor.
XX	US2003124644-A1.
XX	03-JUL-2003.
XX	23-OCT-2001; 2001US-00017471.
XX	23-OCT-2000; 2000US-0242561P.

XX (TAKA)/ TAKANO E.  
PA (BIBB/) BIBB M J.  
XX  
XX Takano E, Bibb MJ;  
DR WPI, 2003-810983/76.  
XX  
PT Modifying an antibiotic-producing strain of Streptomyces coelicolor or  
PT Streptomyces lividans to increase or alter the timing of antibiotic  
PT production in the strain, comprises functionally deleting in the strain  
the scdA or Scdr gene.  
XX  
PS Example 8, Page 9, 3pp; English.

XX This invention relates to the novel modification of an antibiotic-  
CC producing strain of Streptomyces coelicolor or Streptomyces lividans to  
CC increase or to alter the timing of antibiotic production in the strain.  
CC The method comprises functionally deleting in the strain the scdA or Scdr  
CC gene. The method is useful in increasing and altering the timing of  
CC antibiotic production (especially actinohodin and undecylprodigiosin) in  
CC Streptomyces species, particularly Streptomyces coelicolor or  
CC Streptomyces lividans. The present sequence is that of a degenerate PCR  
CC primer which was used for amplification of a region of the Streptomyces  
CC coelicolor ScdA gene during the exemplification of the invention.

SQ Sequence 18 BP; 3 A; 6 C; 5 G; 2 T; 0 U; 2 Other;

Query Match 95.6%; Score 17.2; DB 10; Length 18;  
Best Local Similarity 88.9%; Pred. No. 1.5e+02;  
Matches 16; Conservative 2; Mismatches 0; Indels 0; Gaps 0

OY 1 GACCACTGTCSCGGCATG 18  
|||||.:|:|:  
DB 1 GACCACGTCTCCGCCATG 18

RESULT 3  
ADQ48866  
ID ADQ48866 strand; DNA; 587 BP.  
XX  
XX AC ADQ48866;  
XX  
DT 21-OCT-2004 (first entry)  
XX  
DE Novel canine microarray-related DNA sequence SegID168.  
XX  
XX canine microarray; drug screening; toxicity assay;  
KW environmental pollutant; cellular response; gene expression profile;  
KM toxic response; liver necrosis; fatty liver disease;  
KM protein adduct formation; hepatitis; dog; ds.  
XX  
OS Canis familiaris.  
XX  
PN WO2004063324-A2.  
XX  
PD 29-JUL-2004.  
XX  
PF 05-MAY-2003; 2003WO-US013853.  
XX  
PR 03-MAY-2002; 2002US-0377240P.  
XX  
PA (GENE-) GENE LOGIC INC.  
PA (Pfizer ) PFIZER PROD INC.  
XX  
XX Digians JC, Porter M, Wei T;  
FI  
DR WPI; 2004-561890/54.  
XX  
XX New isolated nucleic acid molecule, useful for drug screening and  
PT toxicity assays or for assessing the impact, including toxicity, of a  
PT compound, pharmaceutical agent or environmental pollutant on a cell or  
living organism.

XX	Claim 1; SEQ ID NO 168; 41bp; English.
PS	
CC	This invention is related to a novel isolated canine nucleic acid
CC	sequences and the construction of canine microarrays containing a
CC	significant portion of the canine genome. The isolated canine nucleic
CC	acid sequences of the invention may be useful for drug screening and
CC	toxicity assays. The invention is therefore useful for assessing the
CC	impact, including toxicity, of a compound, pharmaceutical agent or
CC	environmental pollutant on a cell or living organism. The methods are
CC	useful for detecting genes that are up- or down-regulated in canines in a
CC	disease state. The sequences are useful as diagnostic agents or markers
CC	to detect a cellular response in a sample individually or as part of a
CC	gene expression profile. It is also useful as a target for agents that
CC	modulate gene expression or activity. The database is useful for
CC	producing electronic Northern blots that allow the user to determine the cell
CC	type or tissue in which a given gene is expressed and to allow
CC	determination of the abundance or expression level of a given gene in a
CC	particular tissue or cell. The methods are useful for determining the
CC	similarity of a toxic response to one or more individual compounds. The
CC	methods are useful for predicting at least one toxic response or the
CC	likelihood that a compound or test agent will induce various specific
CC	pathologies such as those of the liver (liver necrosis, fatty liver
CC	disease, protein adduct formation or hepatitis), those of the kidney,
CC	heart, brain or testes, or other pathologies associated with at least one
CC	of the toxins. The methods are also useful for predicting or elucidating
CC	the potential cellular pathways influenced, induced or modulated by the
CC	compound or test agent due to the similarity of the expression profile
CC	compared to the profile induced by a known toxin. The present sequence is
CC	that of a canine DNA sequence which was claimed for use during the
CC	production of a canine microarray of the invention.
XX	
SQ	Sequence 587 BP, 81 A, 186 C, 195 G, 120 T, 0 U, 5 Other;
XX	
Query Match	90.0%; Score 16.2; DB 13; Length 587;
Best Local Similarity	88.2%; Pred. No. 4.1e+02;
Matches	15; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Oy	1 GACCACTGSCCGGCAT 17      ::
Db	401 GACCACGTCCGGGCAT 417
XX	
RESULT 4	
ABN86379/c	
ID	ABN86379 standard; DNA; 4346 BP.
XX	
AC	ABN86379;
XX	
DT	21-OCT-2002 (first entry)
XX	
DE	ScBa, ScBr and ScCb encoding genes containing DNA sequence.
XX	
KW	Antibiotic; bacterium; scba; afsA; scbr; arpa; barA; actinorhodine; Act;
XX	undecylprodigiosin; Red; gene; ds.
OS	Streptomyces coelicolor.
XX	
PN	CA2322241-A1.
XX	
PD	23-APR-2002.
XX	
PF	23-OCT-2000; 2000CA-02322241.
XX	
PR	23-OCT-2000; 2000CA-02322241.
XX	
PA	(PLAN-) PLANT BIOSCIENCE LTD.
XX	
PI	Takano E, Bibb M;
XX	
DR	WPI: 2002-501089/54.
XX	
DR	P-FSDB; ABB80940, ABB80941, ABB80942.
XX	

PT	Modifying antibiotic-producing Streptomyces, to increase, or alter timing
PT	of, antibiotic production, by deleting the scdA or scdr genes.
XX	
PS	Claim 19, 20; Fig 14; 64bp; English.
CC	The invention provides a method for modifying an antibiotic-producing
CC	strain of Streptomyces to increase production of antibiotics or to alter
CC	the timing of antibiotic production. The modification is functional
CC	detection of the scdA gene of S. coelicolor, or its homologues, but is not
CC	detection of the afbA gene of S. griseus, or the modification is
CC	functional deletion of the scdr gene of S. coelicolor, or its homologues,
CC	but is not detection of arpa of S. griseus nor brrA of S. virginiae. The
CC	method is particularly used for production of the antibiotics
CC	actinorhodine (Act) and undecylprodigiosin (Red). The present sequence
CC	represents the nucleic acid sequence of the region containing the S.
CC	coelicolor ScdA, Scdr and Scdb proteins
XX	
SQ	Sequence 4346 BP; 650 A; 1561 C; 1519 G; 616 T; 0 U; 0 Other;
Query Match	88.9%; Score 16; DB 6; Length 4346;
Best Local Similarity	88.9%; Pred. No. 4.7e+02;
Matches 16; Conservative	1; Mismatches 1; Indels 0; Gaps 0
Oy	1 GACCACTGSCCGGCATG 18        :
Db	1453 GACCACGTACCGGCATG 1436
RESULT 5	
ADEI4794/C	
ID	ADEI4794 standard; DNA; 4346 BP.
XX	
AC	ADEI4794;
XX	
DT	29-JAN-2004 (first entry)
XX	
DE	Streptomyces coelicolor partial genome sequence 2.
XX	
KW	antibiotic-producing strain; antibiotic production; scdA gene; Scdr gene;
KM	actinothodin; undecylprodigiosin; gene; ds.
OS	Streptomyces coelicolor.
XX	
FH	Location/Qualifiers
FT	CDS complement(1199..2143)
FT	/tag= a
FT	/product= "Streptomyces coelicolor SrDB protein"
FT	/transl_except= (3149..3147, aa:Tyr)
FT	2261..2908
FT	/tag= b
FT	/product= "Streptomyces coelicolor Srdr protein"
FT	complement(3024..3812)
FT	/tag= c
FT	/product= "Streptomyces coelicolor Srba protein"
FT	/transl_except= (3150..3145, aa:Met)
XX	
PN	US2003124644-A1.
PD	
PD	03-JUL-2003.
XX	
PP	23-OCT-2001; 2001US-00017471.
XX	
PR	23-OCT-2000; 2000US-0242561P.
XX	
PA	(TAKA/) TAKANO E.
PA	(BIBB/) BIBB M J.
XX	
PI	Takano E, Bibb MJ;
XX	
DR	WPI; 2003-810983/76.
DR	P-PsDB; ADEI4791, ADEI4792, ADEI4793.
XX	
PT	Modifying an antibiotic-producing strain of Streptomyces coelicolor or

PT	Streptomyces lividans to increase or alter the timing of antibiotic
PT	production in the strain, comprises functionally deleting in the strain
PT	the scbA or Scbr gene.
XX	
PS	Claim 19; Fig 14; 33pp; English.
XX	
CC	This invention relates to the novel modification of an antibiotic-
CC	producing strain of Streptomyces coelicolor or Streptomyces lividans to
CC	increase or to alter the timing of antibiotic production in the strain.
CC	The method comprises functionally deleting in the strain the scbA or Scbr
CC	gene. The method is useful in increasing and altering the timing of
CC	antibiotic production (especially actinorhodin and undecylprodigiosin) in
CC	Streptomyces species, particularly Streptomyces coelicolor or
CC	Streptomyces lividans. The present sequence is that of a region of the
CC	Streptomyces coelicolor genome, which encodes the scbA, ScbA and Scbr
CC	proteins and which is related to the invention.
XX	
SEQ	Sequence 4346 BP; 650 A; 1561 C; 1519 G; 616 T; 0 U; 0 Other;
Query Match	88.9%; Score 16; DB 10; Length 4346;
Best Local Similarity	88.9%; Pred. No. 4,7e+02;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;	
QY	1 GACCACGTCGCGGCATG 18
Db	1453 GACCACGTCGCGGCATG 1436
RESULT 6	
AACT75505	AACT75505 standard; cDNA; 465 BP.
XX	
AC	AACT75505;
XX	
DT	08-FEB-2001 (first entry)
XX	
DE	Human ORFX ORF1060 polynucleotide sequence SEQ ID NO:2119.
XX	
XX	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX	vulnery; antiporiatic; antiparkinsonian; noctropic; neuroprotective;
XX	anticonvulsant; osteopathic; antiarthritis; immunosuppressant; cardiact;
XX	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX	hypotensive; dermatological; immunosuppressive; antinflammatory;
XX	antiviral; antibacterial; antifungal; antitumoric; antihypertensive;
XX	antineoplastic; gene therapy; cancer; proliferative disorder; hypertension;
XX	neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX	cholesterol ester storage; systemic lupus erythematosus; infection;
XX	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX	bone damage; cartilage damage; antinflammatory disease; coagulation;
XX	thrombosis; contraceptive; ss.
OS	Homo sapiens.
XX	
PN	WO200058473-A2.
XX	
PD	05-OCT-2000.
XX	
PF	31-MAR-2000; 2000WO-US008621.
XX	
PR	31-MAR-1999; 99US-0127607P.
XX	
PR	02-APR-1999; 99US-0127636P.
XX	
PR	05-APR-1999; 99US-0127728P.
XX	
PR	30-MAR-2000; 2000US-00540763.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
PI	Shimkels RA, Leach M;
XX	
DR	WPI; 2000-602362/57.
XX	
DR	P-PSDB; AAB41296.
XX	



PT		Novel nucleic acids and peptides derived from open reading frame X.
PT		Useful for treating e.g. cancers, proliferative disorders,
PT		neurodegenerative disorders and cardiovascular disease.
PS		
XX	Claim 5; Page 1588-1589; 5507pp; English.	
CC	AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,	
CC	which represent the human ORFX open reading frames 1 to 3161. The ORFX	
CC	sequences have activities such as: cytostatic; hepatotropic; vulnerary;	
CC	antiproliferic; antiparkinsonian; nociceptic; neuroprotective; osteopathic;	
CC	anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;	
CC	cardiac; thrombolytic; coagulant; vasotrophic; antidiabetic; hypotensive;	
CC	dermatological; immunosuppressive; antiinflammatory; antibacterial;	
CC	antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The	
CC	sequences can be used for determining the presence of or predisposition	
CC	to, or preventing or treating pathological conditions associated with an	
CC	ORFX-associated disorder. The nucleic acids can be used to express ORFX	
CC	proteins in gene therapy vectors. The proteins and nucleic acids may be	
CC	used to treat cancers, proliferative disorders, neurodegenerative	
CC	disorders, osteoarthritis, graft vs host disease, cardiovascular disease,	
CC	diabetes mellitus, hypertension, hypochloridism, cholesterol ester	
CC	storage, systemic lupus erythematosus, severe combined immunodeficiency	
CC	(SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune	
CC	disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and	
CC	cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to	
CC	enhance coagulation; to inhibit thrombosis; and as a contraceptive	
XX		
SQ	Sequence 465 BP; 102 A; 119 C; 115 G; 128 T; 0 U; 1 Other;	
Query Match	86.7%; Score 15.6; DB 3; Length 465;	
Best Local Similarity	83.3%; Pred. No. 7, 9e+02;	
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0.		
OY	1 GACCACGTCGGCGCATG 18       :  :     106 GACCACGGCCCCGGCATG 123	
Dn		
RESULT 7		
AAX98247		
ID	AAX98247 standard; DNA; 513 BP.	
AC		
XX	AAX98247;	
DT		
XX	25-OCT-1999 (first entry)	
DE		
Nucleotide sequence of ORFC of contig 1344.		
Human pathogen; virulence polypeptide; virulence factor;		
KW	pathogenic infection; Pseudomonas aeruginosa infection; ss.	
XX		
OS	Pseudomonas aeruginosa.	
XX		
PN	WO9927129-A1.	
XX		
PD	03-JUN-1999.	
XX		
PF	25-NOV-1998; 98WO-US0525247.	
XX		
PR	25-NOV-1997; 9TUS-0066517P.	
XX		
PA	(GEHO ) GEN HOSPITAL CORP.	
PI	Ausbrel F, Goodman HM, Rahme LG, Mahajan-Miklos S, Tan M, Cao H;	
PI	Drenkard E, Tsongalis J;	
XX		
DR	WPI; 1999-357851/30.	
XX		
Virulence factors useful in developing disease treatments.		
Disclousure; Fig 32; 228pp; English.		
The present sequence represents a Pseudomonas aeruginosa nucleic acid		

	CC	Sequence P.	aeruginosa	is an opportunistic human pathogen present in soil water and plants. The specification describes virulence polypeptides and nucleic acid sequence encoding such polypeptides. These sequences can be used to identify a compound which is capable of decreasing the expression of a pathogenic virulence factor. Compounds that inhibit the expression or activity of virulence factor polypeptides can be used to treat pathogenic infections, especially where the infection is a P. aeruginosa infection. note: the sequences given in the specification were poorly legible, and in some instances assumptions were made as to the identity of the base; it is therefore possible that the sequence given below is not entirely correct
Dd	Db	Query Match Best Local Similarity Matches    15;	86.7%; 83.3%; Conservative     2;	Mismatches       1;  Indels            0;  Gaps              0
Oy		1 GACCACTGTCCSCGGCATG 18        :::		
		133 GACCAGTTCGCCGCATG 150		
RESULT 8				
ADR91782/C	ID	ADR91782 standard; DNA; 612 BP.		
AC	XX	ADR91782;		
DT	DT	16-DEC-2004 (first entry)		
DE	XX	Novel S. pneumoniae DNA sequence, SEQ ID 417.		
KW	XX	Meningitis; bacteraemia; pneumonia; otitis media; ds;		
XN	XX	bacterial infection.		
OS	XX	Streptococcus pneumoniae.		
PB	PN	US6800744-BI.		
PD	PP	05-OCT-2004.		
PR	XX	30-JUN-1998;    98US-00107433.		
PS	PF	02-JUL-1997;    97US-0051553P.		
RA	PR	12-MAY-1998;    98US-0085131P.		
SQ	XX	(GENO-) GENOME THERAPEUTICS CORP.		
TU	PA	Doucette-Stamm LA, Bush D;		
VZ	PI	WI; 2004-697205/68.		
ZD	DR	P-PSTDB; ADR94385.		
	PT	New isolated nucleic acid encoding a Streptococcus pneumoniae polypeptide, useful for diagnosing, preventing and/or treating pathological conditions resulting from the bacterial infection.		
	PS	Disclosure; SEQ ID NO 417; 151bp; English.		
	XX	The invention relates to an isolated nucleic acid comprising a sequence encoding a Streptococcus pneumoniae ADR9136polypeptide, or its fragments, with any of 9 fully defined sequences (appearing as ADR94308,		
	CC	ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,		
	CC	ADR96079) or any of the fully defined sequences appearing as ADR91705,		
	CC	ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or		
	CC	ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide		
	CC	sequences, or at least 40, 60 or 300 consecutive nucleotides, which is		
	CC	hybridisable under high stringency conditions to the nucleotide sequence.		
	CC	The nucleic acids and proteins are chosen from 5206 disclosed sequences.		
	CC	Also included are a recombinant expression vector comprising the isolated		
	CC	nucleic acid cited above operably linked to a transcription regulatory		
	CC	element, a cell comprising the recombinant expression vector and a probe		



CC comprising at least 20 consecutive nucleotides of the nucleotide  
 CC sequences as cited above. The methods and compositions of the present  
 CC invention are useful for the diagnosis, prevention and/or treatment of  
 CC pathological conditions resulting from bacterial infection by  
 CC *Streptococcus pneumoniae* e.g. pneumonia, bacteraemia, meningitis and  
 CC otitis media. The present sequence is one of the 2603 disclosed S.  
 CC *pneumoniae* nucleic acid sequences. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from USPTO at  
 CC [seqdata.uspto.gov/sequence.html?docID=680074481](http://seqdata.uspto.gov/sequence.html?docID=680074481).

XX Sequence 612 BP; 164 A; 122 C; 135 G; 191 T; 0 U; 0 Other;

Query Match 86.7%; Score 15.6; DB 13; Length 612;

Best Local Similarity 83.3%; Pred. No. 7.8e+02;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGTCGCGCATG 18  
 |||||:|||||  
 Db 160 GACCACTGTCGCGCATG 143

# RESULT 9

ID ABL10881 standard; cDNA; 744 BP.

XX ABL10881;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 27125.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX MO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009221.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX P-PSDB; ABB66778.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 XX genes from Drosophila and for elucidating cell signalling and cell-cell  
 XX interactions.

XX Claim 1; SEQ ID NO 27125; 21pp + Sequence listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB16177), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins (AB57737-  
 CC AB572072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)

XX Sequence 744 BP; 174 A; 249 C; 192 G; 129 T; 0 U; 0 Other;

Query Match 86.7%; Score 15.6; DB 4; Length 744;

Best Local Similarity 83.3%; Pred. No. 7.8e+02;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACTGTCGCGCATG 18  
 |||||:|||||  
 Db 454 GTCACGTCGCGCATG 471

# RESULT 10

ID AAX98245/c  
 AAX98245 standard; DNA; 766 BP.

XX AAX98245;

XX 25-OCT-1999 (first entry)

XX Nucleotide sequence of ORF4 of contig 1344.

XX Human pathogen; virulence polypeptide; virulence factor;

XX pathogenic infection; *Pseudomonas aeruginosa* infection; ss.

XX *Pseudomonas aeruginosa*.

XX Key Location/Qualifiers

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

XX misc\_feature

/\*tag= a  
 /note= "these bases represent nucleotides missing from  
 the sequence given in the specification; They are  
 included to maintain the nucleotide numbering given in  
 the specification for this sequence"  
 641..650  
 /\*tag= b  
 /note= "these bases represent nucleotides missing from  
 the sequence given in the specification; They are  
 included to maintain the nucleotide numbering given in  
 the specification for this sequence"  
 681..700  
 /\*tag= c  
 /note= "these bases represent nucleotides missing from  
 the sequence given in the specification; They are  
 included to maintain the nucleotide numbering given in  
 the specification for this sequence"  
 721..730  
 /\*tag= d  
 /note= "these bases represent nucleotides missing from  
 the sequence given in the specification; They are  
 included to maintain the nucleotide numbering given in  
 the specification for this sequence"  
 761..766  
 /\*tag= e  
 /note= "these bases represent nucleotides missing from  
 the sequence given in the specification; They are  
 included to maintain the nucleotide numbering given in  
 the specification for this sequence"

XX MO9927129-A1.

XX 03-JUN-1999.

XX 25-NOV-1998; 98WO-US025247.

XX 25-NOV-1997; 97US-0066517P.

XX (GENO ) GEN HOSPITAL CORP.

XX Ausubel F, Goodman HM, Rahne LG, Mahajan-Miklos S, Tan M, Cao H;

XX Drenkard E, Tsongalis J;

XX WPI; 1999-357851/30.

XX Virulence factors useful in developing disease treatments.

XX Disclosure; Fig 32; 228pp; English.

of *Pseudomonas* species using biochip technology. Sequences ABD01397-CC ABD17967 represent *P. aeruginosa* polynucleotides of the invention. Note: CC The sequence data for this patent did not form part of the printed CC specification but was obtained in electronic format from USPTO at CC [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html) CC

50 Sequence 891 BP; 130 A; 286 C; 308 G; 167 T; 0 U; 0 Other;

Query Match 86.7%; Score 15.6; DB 11; Length 891;  
Best Local Similarity 83.3%; Pred. No. 7.7e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0

1 GACCACGTSCCGGCATG 18  
|||||||:|||||  
Db 432 GACCACGTCCGCGCATG 415

RESULT 12  
ABX05949/C  
ID ABX05949 standard; DNA; 1188 BP.  
XX ABX05949;  
AC  
XX  
XX 27-OCT-2003 (revised)  
DT 11-FEB-2003 (first entry)  
XX  
XX  
DE 5. pneumoniae type 4 strain coding region #237.  
XX  
XX Gene; ds: bacterial meningitis; pneumonia; sepsis; otitis media;  
KM ear infection; antiinflammatory; antibacterial; immunostimulant;  
XX audiotory; respiratory; gene therapy; vaccine.  
XX  
XX  
OS Streptococcus pneumoniae; type 4 strain.  
XX  
XX WO200277021-A2.  
PM 03-OCT-2002.  
PD  
XX  
XX 27-MAR-2002; 2002WO-IB02163.  
PF  
XX  
XX 27-MAR-2001; 2001GB-00007658.  
PR  
XX  
XX (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
XX Masignani V, Tettelin H, Frazer C;  
PI WPI; 2003-040579/03.  
DR P-PSDB; ABU00670.  
XX  
XX  
PT New proteins and nucleic acid molecules from *Streptococcus pneumoniae*,  
PT useful as medicaments for treating or preventing a disease or infection  
PT due to *Streptococcus bacteria*, such as pneumonia, sepsis, otitis media or  
PT ear infection.  
XX  
XX  
PS Claim 6; SEQ ID NO 473; 56bp; English.  
XX  
XX  
CC The invention relates to a protein comprising or having at least 50%  
CC identity to any of the 2469 amino acid sequences, identified in the  
CC specification (available on a computer readable format), or its fragment.  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC *Streptococcus pneumoniae* type 4 strain genomic sequence appearing as  
CC AB556654. Also included are an antibody which binds one of the proteins,  
CC treating a patient by administering the protein, DNA or antibody (in a  
CC composition), a kit comprising first and second primers, which are a  
CC nucleic acid cited above or fragments between nucleotides 8-100 of the  
CC sequence not defined in the specification, for amplifying a target  
CC sequence contained within a *Streptococcus* nucleic acid sequence, where  
CC the first primer is substantially complementary to the target sequence  
CC and the second primer is substantially complementary to the complement of  
CC the target sequence, and where the parts of the primers having  
CC substantial complementarity define the termini of the target sequence to  
CC be amplified, assay comprising contacting a test compound with the

of *Pseudomonas* species using biochip technology. Sequences ABD01397-CC ABD17967 represent *P. aeruginosa* polynucleotides of the invention. Note: CC The sequence data for this patent did not form part of the printed CC specification but was obtained in electronic format from USPTO at CC [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html) CC

50 Sequence 891 BP; 130 A; 286 C; 308 G; 167 T; 0 U; 0 Other;

Query Match 86.7%; Score 15.6; DB 11; Length 891;  
Best Local Similarity 83.3%; Pred. No. 7.7e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0

1 GACCACGTSCCGGCATG 18  
|||||||:|||||  
Db 432 GACCACGTCCGCGCATG 415

RESULT 12  
ABX05949/C  
ID ABX05949 standard; DNA; 1188 BP.  
XX  
XX ABX05949;  
AC  
XX  
XX 27-OCT-2003 (revised)  
DT 11-FEB-2003 (first entry)  
XX  
XX  
DE 5. pneumoniae type 4 strain coding region #237.  
XX  
XX Gene; ds: bacterial meningitis; pneumonia; sepsis; otitis media;  
KM ear infection; antiinflammatory; antibacterial; immunostimulant;  
XX audiotory; respiratory; gene therapy; vaccine.  
XX  
XX  
OS Streptococcus pneumoniae; type 4 strain.  
XX  
XX WO200277021-A2.  
PM  
PD 03-OCT-2002.  
XX  
XX 27-MAR-2002; 2002WO-IB002163.  
PF  
XX 27-MAR-2001; 2001GB-00007658.  
PR  
XX  
XX (CHIR-) CHIRON SPA.  
PA (GENO-) INST GENOMIC RES.  
XX  
XX Masignani V, Tettelin H, Frazer C;  
PI WPI; 2003-040579/03.  
XX  
XX P-PSDB; ABU00670.  
DR  
XX  
XX New proteins and nucleic acid molecules from *Streptococcus pneumoniae*,  
PT useful as medicaments for treating or preventing a disease or infection  
PT due to *Streptococcus bacteria*, such as pneumonia, sepsis, otitis media or  
PT ear infection.  
XX  
XX  
XX Claim 6; SEQ ID NO 473; 56bp; English.  
PS  
XX  
XX The invention relates to a protein comprising or having at least 50%  
CC identity to any of the 2469 amino acid sequences, identified in the  
CC specification (available on a computer readable format), or its fragment.  
CC expressed from 2469 of 2489 identified DNA coding regions from the  
CC *Streptococcus pneumoniae* type 4 strain genomic sequence appearing as  
CC AB556654. Also included are an antibody which binds one of the proteins,  
CC treating a patient by administering the protein, DNA or antibody (in a  
CC composition), a kit comprising first and second primers, which are a  
CC nucleic acid cited above or fragments between nucleotides 8-100 of the  
CC sequence not defined in the specification, for amplifying a target  
CC sequence contained within a *Streptococcus* nucleic acid sequence, where  
CC the first primer is substantially complementary to the target sequence  
CC and the second primer is substantially complementary to the complement of  
CC the target sequence, and where the parts of the primers having  
CC substantial complementarity define the termini of the target sequence to  
CC be amplified, assay comprising contacting a test compound with the

CC protein, and determining whether the test compound binds to the protein  
CC and a Streptococcus pneumoniae bacterium, where one or more genes  
CC encoding the proteins has been rendered inactive. The proteins, nucleic  
CC acid molecules, antibody and compositions are useful as medicaments for  
CC treating or preventing a disease or infection due to streptococcus  
CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis  
CC media or ear infection. They are also useful in developing vaccines,  
CC diagnostics and antibiotics. The methods are useful for identifying  
CC immunodominant proteins. The present sequence is one of the 2489  
CC identified coding region from the genomic sequence. Note: The sequence  
CC data for this patent did not form part of the printed specification, but  
CC was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 27-OCT-2003 to  
CC standardise OS field)

XX  
SQ Sequence 1188 BP; 355 A; 228 C; 261 G; 344 T; 0 U; 0 Other;

QY Query Match 86.7%; Score 15.6; DB 10; Length 1188;  
Best Local Similarity 83.3%; Pred. No. 7.7e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1 GACCACTGCTCCGCGCATG 18  
739 GACCACTGCTCCGCGCATG 722

XX  
RESULT 13  
ABD15034/c  
ID ABD15034 standard; DNA; 1275 BP.

XX  
AC ABD15034;  
XX  
DT 29-JUL-2004 (first entry)

XX  
DE Pseudomonas aeruginosa polynucleotide #13638.

XX  
KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
XX antibacterial.

XX  
OS Pseudomonas aeruginosa.

XX  
PN US6551795-B1.

XX  
PD 22-APR-2003.

XX  
PF 18-FEB-1999; 99US-00252991.

XX  
PR 18-FEB-1998; 98US-0074788P.  
XX 27-JUL-1998; 98US-0094190P.

XX  
PA (GENO-) GENOME THERAPEUTICS CORP.

XX  
PI Rubenfeld MJ, Nolling J, Deloughery C, Bush D;  
XX P-PSDB; AB081463.

XX  
DR WPI; 2003-615309/58.

XX  
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
XX useful as molecular targets for diagnostics, prophylaxis and treatment of  
XX pathological conditions resulting from bacterial infection.

XX  
PS Disclosure; SEQ ID NO 13638; 455bp; English.

XX  
CC The invention relates to Pseudomonas aeruginosa polypeptides and the  
XX polynucleotides encoding them. The sequences are useful in diagnosis and  
XX therapy of pathological conditions, as molecular targets for diagnostics,  
XX prophylaxis and treatment of pathological conditions resulting from a  
XX bacterial infection, for evaluating a compound, such as a polypeptide,  
XX for the ability to bind a P. aeruginosa nucleic acid, as components of  
XX effective antibacterial targets, as targets for antibacterial drugs,  
XX including anti-P. aeruginosa drugs, as templates for recombinant  
XX production of P. aeruginosa-derived peptides or polypeptides, as target  
XX components for diagnosis and/or treatment of P. aeruginosa-caused

CC infection, and in detection of P. aeruginosa sequences or other sequences  
CC of Pseudomonas species using biotin technology. Sequences ABD01397-  
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification but was obtained in electronic format from USPTO at  
CC segdata.uspto.gov/sequence.html

XX  
SQ Sequence 1275 BP; 208 A; 409 C; 443 G; 215 T; 0 U; 0 Other;

QY Query Match 86.7%; Score 15.6; DB 11; Length 1275;  
Best Local Similarity 83.3%; Pred. No. 7.6e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1 GACCACTGCTCCGCGCATG 18  
367 GACCACTGCTCCGCGCATG 350

XX  
RESULT 14  
AAZ32291  
ID AAZ32291 standard; DNA; 1408 BP.

XX  
AC AAZ32291;  
XX  
DT 20-JAN-2000 (first entry)

XX  
DE Sphingomonas sp. biotin synthase biob nucleotide sequence.

XX  
KW Sphingomonas; biotin; biotin synthase; desbiobiotin; fermentation; ds.

XX  
OS Sphingomonas sp.

XX  
PN JP11276175-A.

XX  
PD 12-OCT-1999.

XX  
PF 31-MAR-1998; 98JP-00086972.

XX  
PR 31-MAR-1998; 98JP-00086972.

XX  
PA (SUMO) SUMITOMO CHEM CO LTD.

XX  
DR WPI; 1999-626939/54.

XX  
PT Preparation of biotin - comprising fermentation using Escherichia coli.  
XX Example; Page 11-12; 13pp; Japanese.

XX  
CC A method has been developed for the preparation of biotin by fermentation  
XX using a microbe in which Escherichia coli having a plasmid containing a  
XX gene coding an enzyme having biotin synthase activity is added to the  
XX culture liquid. The method can convert desbiobiotin accumulated in the  
XX culture liquid to biotin efficiently. The present sequence represents a  
XX exemplification of the present invention

XX  
SQ Sequence 1408 BP; 214 A; 540 C; 442 G; 212 T; 0 U; 0 Other;

QY Query Match 86.7%; Score 15.6; DB 2; Length 1408;  
Best Local Similarity 83.3%; Pred. No. 7.6e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Db 1 GACCACTGCTCCGCGCATG 18  
1057 GACCACTGCTCCGCGCATG 1074

XX  
RESULT 15  
AAZ32287  
ID AAZ32287 standard; DNA; 1408 BP.

XX  
AC AAZ32287;  
XX

```
DT 20-JAN-2000 (first entry)
XX Sphingomonas sp. biotin synthase biof nucleotide sequence.
DE Sphingomonas; biotin; biotin synthase; desbiobiotin; fermentation; ds.
XX Sphingomonas sp.
OS Sphingomonas sp.
XX Jp11276175-A.
PN 12-OCT-1999.
PD 31-MAR-1998; 98JP-00086972.
XX 31-MAR-1998; 98JP-00086972.
XX 31-MAR-1998; 98JP-00086972.
XX (SUMO) SUMITOMO CHEM CO LTD.
XX WPI, 1999-626939/54.
DR WPI, 1999-626939/54.
XX Preparation of biotin - comprising fermentation using Escherichia coli.
PT Example; Page 9-10; 13pp; Japanese.
PS Example; Page 9-10; 13pp; Japanese.
XX A method has been developed for the preparation of biotin by fermentation
CC using a microbe in which Escherichia coli having a plasmid containing a
CC gene coding an enzyme having biotin synthase activity is added to the
CC culture liquid. The method can convert desbiobiotin accumulated in the
CC culture liquid to biotin efficiently. The present sequence represents a
CC Sphingomonas sp. gene having biotin synthase activity, used in the
CC exemplification of the present invention.
XX Sequence 1408 BP; 214 A; 541 C; 441 G; 212 T; 0 U; 0 Other;
SQ Query Match 86.7%; Score 15.6; DB 2; Length 1408;
Best Local Similarity 83.3%; Pred. No. 7.6e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Oy 1 GACCAAGTSCGCGCATG 18
Db 1057 GACCTCGTCCGCGCATG 1074
RESULT 16
ADA73004
XX ADA73004 standard; DNA; 2000 BP.
XX ADA73004;
AC 20-NOV-2003 (first entry)
XX 20-NOV-2003 (first entry)
DT Rice gene, SEQ ID 6330.
XX Rice gene, SEQ ID 6330.
DE Plant; bacterial infection; fungal infection; viral infection; rice;
XX gene; ds.
XX Oryza sativa.
OS Oryza sativa.
XX WO2003000898-A1.
PN WO2003000898-A1.
XX 03-JAN-2003.
PD 22-JUN-2001; 2001WO-1B001105.
XX 22-JUN-2001; 2001WO-1B001105.
PR 22-JUN-2001; 2001WO-1B001105.
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX WPI, 2003-175290/17.
XX
```

```
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX Claim 27; SEQ ID NO 6330; 899pp; English.
PS Claim 27; SEQ ID NO 6330; 899pp; English.
XX The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX Sequence 2000 BP; 419 A; 529 C; 502 G; 547 T; 0 U; 3 Other;
SQ Query Match 86.7%; Score 15.6; DB 8; Length 2000;
Best Local Similarity 83.3%; Pred. No. 7.5e+02;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Oy 1 GACCAAGTSCGCGCATG 18
Db 1089 GGCACGTGCCGCGCATG 1106
RESULT 17
AA98244
XX AA98244 standard; DNA; 2048 BP.
XX AA98244;
AC 25-OCT-1999 (first entry)
DT 25-OCT-1999 (first entry)
XX 25-OCT-1999 (first entry)
DE Contig 1344 identified using sequence tag 33C7.
XX Contig 1344 identified using sequence tag 33C7.
XX Human pathogen; virulence polypeptide; virulence factor;
XX pathogenic infection; Pseudomonas aeruginosa infection; ss.
XX Pseudomonas aeruginosa.
OS Pseudomonas aeruginosa.
XX Key
XX Location/Qualifiers
FH 1141..1444
FT misc_feature
FT /tag= a
FT /note= "these bases represent nucleotides missing from
FT the sequence given in the specification; They are
FT included to maintain the nucleotide numbering given in
FT the specification for this sequence"
FT 1201..1210
FT misc_feature
FT /tag= b
FT /note= "these bases represent nucleotides missing from
FT the sequence given in the specification; They are
FT included to maintain the nucleotide numbering given in
FT the specification for this sequence"
FT 1261..1290
FT misc_feature
FT /tag= c
FT /note= "these bases represent nucleotides missing from
FT the sequence given in the specification; They are
FT included to maintain the nucleotide numbering given in
FT the specification for this sequence"
FT 1321..1350
FT misc_feature
FT /tag= d
FT /note= "these bases represent nucleotides missing from
FT the sequence given in the specification; They are
FT included to maintain the nucleotide numbering given in
FT the specification for this sequence"
FT 1371..1400
FT misc_feature
FT /tag= e
FT /note= "these bases represent nucleotides missing from
FT the sequence given in the specification; They are
```

FT included to maintain the nucleotide numbering given in  
 FT the specification for this sequence"  
 FT 1431..1460  
 FT /\*tag= f  
 FT /note= "these bases represent nucleotides missing from  
 FT the sequence given in the specification; they are  
 FT included to maintain the nucleotide numbering given in  
 FT the specification for this sequence"  
 XX  
 XX WO9927129-A1.  
 XX  
 XX 03-JUN-1999.  
 XX  
 XX 25-NOV-1998; 98WO-US025247.  
 XX  
 XX 25-NOV-1997; 97US-006517P.  
 XX  
 XX (GEHO ) GEN HOSPITAL CORP.  
 XX  
 XX Ausubel F, Goodman HW, Rahne LG, Mahajan-Miklos S, Tan M, Cao H,  
 XX Drenkard E, Tsongalis J;  
 XX WPI, 1999-357851/30.  
 XX  
 XX virulence factors useful in developing disease treatments.  
 XX  
 XX Disclosure; Fig 32; 228pp; English.  
 XX  
 XX The present sequence represents a Pseudomonas aeruginosa nucleic acid  
 XX sequence. P. aeruginosa is an opportunistic human pathogen present in  
 XX soil water and plants. The specification describes virulence polypeptides  
 XX and nucleic acid sequence encoding such polypeptides. These sequences can  
 XX be used to identify a compound which is capable of decreasing the  
 XX expression of a pathogenic virulence factor. Compounds that inhibit the  
 XX expression or activity of virulence factor polypeptides can be used to  
 XX treat pathogenic infections, especially where the infection is a P.  
 XX aeruginosa infection. note: the sequences given in the specification were  
 XX poorly legible, and in some instances assumptions were made as to the  
 XX identity of the base; it is therefore possible that the sequence given  
 XX below is not entirely correct  
 XX  
 XX Sequence 2048 BP; 322 A; 676 C; 594 G; 322 T; 0 U; 134 Other;  
 XX  
 XX Query Match 86.7%; Score 15.6; DB 2; Length 2048;  
 XX Best Local Similarity 83.3%; Pred. No. 7.5e+02;  
 XX Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 XX  
 XX 1 GACCACTGCTCCGGCGCATG 18  
 XX |||||:|:|:|:|  
 XX Db 1069 GACCACTGCTCCGGCGCATG 1086  
 XX  
 XX RESULT 18  
 XX AAV52338/c  
 XX ID AAV52338 standard; DNA; 2395 BP.  
 XX  
 XX AAV52338;  
 XX  
 XX 23-OCT-1998 (first entry)  
 XX  
 XX Streptococcus pneumoniae genome fragment SEQ ID NO:205.  
 XX  
 XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;  
 XX computer readable medium; vaccine; pharmaceutical composition; ds.  
 XX  
 XX Streptococcus pneumoniae.  
 XX  
 XX WO9818931-A2.  
 XX  
 XX 07-MAY-1998.  
 XX  
 XX 30-OCT-1997; 97WO-US019588.  
 XX

PR 31-OCT-1996; 96US-0029960P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Kunsch CA, Choi GH, Dillon PJ, Rosen CA, Barash SC, Fannon M;  
 XX Dougherty BA;  
 XX WPI, 1998-272225/24.  
 XX  
 XX Computer-readable medium with recorded Streptococcus pneumoniae  
 XX polynucleotide sequences - useful in diagnostic kits and assays, and  
 XX pharmaceutical compositions and vaccines for Streptococcus pneumoniae.  
 XX  
 XX Claim 1; Page 1189-1191; 1409pp; English.  
 XX  
 XX The present invention describes a computer readable medium which has the  
 XX nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524) recorded  
 XX on it, or a representative fragment or a sequence at least 95% identical  
 XX to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1  
 XX to 391 (AAV52134 to AAV52524) are genomic fragments from Streptococcus  
 XX pneumoniae. The present invention also describes an isolated nucleic acid  
 XX molecule encoding a homologue of any of the fragments of the S. pneumoniae  
 XX genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced  
 XX by a process comprising: (a) screening a genomic DNA library using as a  
 XX probe a target sequence defined by any of the sequences in SEQ ID NO:1 to  
 XX 391, identifying members of the library which contain sequences that  
 XX hybridize to the target sequence and isolating the nucleic acid molecules  
 XX from the members; or (b) isolating mRNA, DNA or cDNA produced from an  
 XX organism, amplifying nucleic acid molecules whose nucleotide sequence is  
 XX homologous to amplification primers derived from the fragment of the S.  
 XX pneumoniae genome to prime the amplification and isolating the amplified  
 XX sequences. The computer readable medium can be used in a computer-based  
 XX system for identifying fragments of the S. pneumoniae genome of  
 XX commercial importance, or expression modulating fragments of the S.  
 XX pneumoniae genome. Products from the present invention can be used in  
 XX diagnosis kits and assays, and pharmaceutical compositions and vaccines  
 XX for S. pneumoniae  
 XX  
 XX Sequence 2395 BP; 755 A; 415 C; 548 G; 677 T; 0 U; 0 Other;  
 XX  
 XX Query Match 86.7%; Score 15.6; DB 2; Length 2395;  
 XX Best Local Similarity 83.3%; Pred. No. 7.5e+02;  
 XX Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 XX  
 XX 1 GACCACTGCTCCGGCGCATG 18  
 XX |||||:|:|:|:|  
 XX Db 1257 GACCACTGCTCCGGCGCATG 1240  
 XX  
 XX RESULT 19  
 XX AEN59696  
 XX ID AEN59696 standard; cDNA; 5042 BP.  
 XX  
 XX AEN59696;  
 XX  
 XX 28-JUN-2002 (first entry)  
 XX  
 XX Novel human coding sequence SEQ ID NO: 107.  
 XX  
 XX Human; anti-infective; vaccine; anti-inflammatory; immunomodulator;  
 XX human; anti-infective; vaccine; anti-inflammatory; immunomodulator;  
 XX anti-infective; vaccine; anti-inflammatory; immunomodulator;  
 XX neuroprotective; anti-infective; vaccine; anti-inflammatory; immunomodulator;  
 XX expressed sequence tag; gene; ss.  
 XX  
 XX Homo sapiens.  
 XX  
 XX WO200222660-A2.  
 XX  
 XX 21-MAR-2002.  
 XX  
 XX 10-SEP-2001; 2001WO-US026015.  
 XX  
 XX 11-SEP-2000; 2000US-00659671.  
 XX

XX (HYSE-) HYSEQ INC.  
PA Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT,  
XX WPI; 2002-292408/33.  
DR P-PSDB; ABB97283.  
XX An isolated polynucleotide for treating diseases associated with its  
PT encoded polypeptide such as cancer and multiple sclerosis.  
XX Claim 1; SEQ ID NO 107; 509pp; English.  
XX The present invention provides the protein and coding sequences of 444  
CC novel human proteins. These were isolated from expressed sequences tags  
CC (ESTs). They can be used to stimulate cell growth, to regulate  
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat  
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke  
CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.  
CC rheumatoid arthritis, and to treat nervous system disorders e.g.  
CC Parkinson's disease. The present sequence is a coding sequence of the  
CC invention  
XX  
SQ Sequence 5042 BP; 1448 A; 1034 C; 1199 G; 1361 T; 0 U; 0 Other;  
Query Match 86.7%; Score 15.6; DB 6; Length 5042;  
Best Local Similarity 83.3%; Pred. No. 7.3e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GACCACTGTCGCGCATG 18  
DB 149 GACCACTGTCGCGCATG 166  
RESULT 20  
ADE06975 ADE06975 standard; DNA; 5362 BP.  
XX ADE06975;  
AC ADE06975;  
XX 29-JAN-2004 (first entry)  
DT  
XX Novel coding sequence (useful for identifying genetic disorders) #41.  
XX novel gene; novel protein; tissue marker; molecular weight marker;  
KM chromosome marker; genetic disorder; gene; ds.  
XX Unidentified.  
OS  
XX WO2003054152-A2.  
PN  
XX 03-JUL-2003.  
PD  
XX 10-DEC-2002; 2002WO-US039555.  
PF  
XX 10-DEC-2001; 2001US-0339739P.  
PR 11-DEC-2001; 2001US-0339453P.  
PR 14-MAR-2002; 2002US-0365091P.  
PR 14-MAR-2002; 2002US-0365384P.  
PR 12-APR-2002; 2002US-0372381P.  
PR 12-APR-2002; 2002US-0372615P.  
PR 22-APR-2002; 2002US-0012855P.  
PR 24-APR-2002; 2002US-0376045P.  
XX  
XX (HYSE-) HYSEQ INC.  
PA Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;  
PI Ghosh W, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;  
PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ,  
XX

DR WPI; 2003-569235/53.  
DR P-PSDB; ADE07886C.  
XX  
XX New polynucleotides, useful for expressing recombinant proteins for  
PT analysis, characterization or therapeutic use, or as markers for tissues  
PT in which the corresponding protein is preferentially expressed.  
XX  
XX Claim 1; SEQ ID NO 41; 1177pp; English.  
PS  
XX The invention comprises the amino acid and coding sequences of novel  
CC proteins. The DNA and protein sequences of the invention are useful as:  
CC markers for tissues in which the corresponding protein is preferentially  
CC expressed; as molecular weight markers on gels; as chromosome markers or  
CC tags; to identify chromosomes or to map related gene positions; and to  
CC compare with endogenous DNA sequences in patients to identify potential  
CC genetic disorders. The present DNA sequence represents a gene of the  
CC invention.  
XX  
SQ Sequence 5362 BP; 1522 A; 1122 C; 1278 G; 1440 T; 0 U; 0 Other;  
Query Match 86.7%; Score 15.6; DB 10; Length 5362;  
Best Local Similarity 83.3%; Pred. No. 7.3e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GACCACTGTCGCGCATG 18  
DB 70 GACCACTGTCGCGCATG 87  
RESULT 21  
AAH89920 AAH89920 standard; cDNA; 5624 BP.  
XX AAH89920;  
AC AAH89920;  
XX 01-OCT-2001 (first entry)  
DT  
XX Human bone marrow cDNA, SEQ ID NO: 51.  
DE  
XX Human; bone marrow; antiinflammatory; cytostatic; neuroprotective;  
KM antiviral; antibacterial; antifungal; anti-HIV; haemostatic;  
KM immunosuppressive; gene therapy; cytokine cell proliferation;  
KM cell differentiation modulator; immune disorder; infection; cancer;  
KM human immunodeficiency virus; HIV; autoimmune disorder; haemophilia; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200153453-A2.  
PN  
XX 26-JUL-2001.  
PD  
XX 23-DEC-2000; 2000WO-US034960.  
PF  
XX 23-DEC-1999; 99US-00481275.  
PR 21-JAN-2000; 2000US-00488725.  
PR 25-APR-2000; 2000US-0052317.  
PR 20-JUL-2000; 2000US-00598042.  
PR 19-JUL-2000; 2000US-00620312.  
PR 03-AUG-2000; 2000US-00653450.  
PR 14-SEP-2000; 2000US-00662191.  
PR 19-OCT-2000; 2000US-00693036.  
PR 30-NOV-2000; 2000US-0250583P.  
XX  
XX (HYSE-) HYSEQ INC.  
PA Ford UE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;  
PI Ren F, Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;  
PI Zhou P, Drmanac RT;  
XX  
XX WPI; 2001-488707/53.  
DR P-PSDB; AAM00801.  
XX Novel bone-marrow-expressed polynucleotides and polypeptides, useful for

PT treating e.g. cancer and immune deficiency disorders.  
XX  
PS Claim 1, Page 238-240, 649pp, English.  
XX  
CC The present sequence is one of 251 novel human polynucleotides expressed in the bone marrow. The polynucleotide and the polypeptide encoded by it are useful in the treatment of various immune deficiencies and disorders. The deficiencies and disorders may be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal infection, or may result from an autoimmune disorder, a coagulation disorder (e.g. haemophilia), inhibition of tumour cell proliferation, suppression of an inflammatory response or treatment of a nervous system disorder such as Alzheimer's disease. Detection of the presence or increased expression of the polynucleotide or the protein it encodes is useful for the diagnosis and/or prognosis of one or more types of cancer. The polynucleotide and polypeptide can be used as nutritional sources or supplements and in the screening of chemical compounds as potential drugs  
CC  
XX Sequence 5624 BP, 1555 A; 1205 C; 1349 G; 1515 T; 0 U; 0 Other;  
SQ  
Query Match 86.7%; Score 15.6; DB 5; Length 5624;  
Best Local Similarity 83.3%; Pred. No. 7.2e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 GACCACTGCGCGCATG 18  
Db 85 GACCACTGCGCGCATG 102  
RESULT 22  
ABLI0880/c  
ID ABLI0880 standard; cDNA; 12332 BP.  
XX  
AC ABLI0880;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 27122.  
XX  
KM Drosophila; developmental biology; cell signalling; insecticide;  
XX  
KW pharmaceutical; gene; ds.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
DR P-PSDB; ABB66777.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.  
PT  
XX  
PS Claim 1, SEQ ID NO 27122, 21pp + Sequence Listing, English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABLI6176-ABLI0511), expressed DNA sequences (ABLI01840-ABLI6175) and the encoded proteins (ABB57737-

CC ABB72072). The sequence data for this patent did not form part of the CC printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/publicated\\_pct\\_sequences](http://ftp.wipo.int/pub/publicated_pct_sequences)  
XX  
SQ Sequence 12332 BP, 3452 A; 2580 C; 2685 G; 3615 T; 0 U; 0 Other;  
Qy 1 GACCACTGCGCGCATG 18  
Db 1291 GTCCACTGCGCGCATG 1274  
RESULT 23  
ABSS6454 01/c  
Continuation (2 of 22) of ABSS6454 from base 100001 (Streptococcus pneumoniae type 4 str.  
WP Sequence split into 22 fragments LOCUS ABSS6454 Accession ABSS6454  
WP Fragment Name Begin End  
WP ABSS6454\_00 1 110000  
WP ABSS6454\_01 100001 210000  
WP ABSS6454\_02 200001 310000  
WP ABSS6454\_03 300001 410000  
WP ABSS6454\_04 400001 510000  
WP ABSS6454\_05 500001 610000  
WP ABSS6454\_06 600001 710000  
WP ABSS6454\_07 700001 810000  
WP ABSS6454\_08 800001 910000  
WP ABSS6454\_09 900001 1010000  
WP ABSS6454\_10 1000001 1110000  
WP ABSS6454\_11 1100001 1210000  
WP ABSS6454\_12 1200001 1310000  
WP ABSS6454\_13 1300001 1410000  
WP ABSS6454\_14 1400001 1510000  
WP ABSS6454\_15 1500001 1610000  
WP ABSS6454\_16 1600001 1710000  
WP ABSS6454\_17 1700001 1810000  
WP ABSS6454\_18 1800001 1910000  
WP ABSS6454\_19 1900001 2010000  
WP ABSS6454\_20 2000001 2110000  
WP ABSS6454\_21 2100001 2162598  
Query Match 86.7%; Score 15.6; DB 10; Length 110000;  
Best Local Similarity 83.3%; Pred. No. 6.5e+02;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 GACCACTGCGCGCATG 18  
Db 90126 GACCACTGCGCGCATG 90109  
RESULT 24  
ABLI6420/c  
ID ABLI6420 standard; DNA; 5356 BP.  
XX  
AC ABLI6420;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 733.  
XX  
KM Drosophila; developmental biology; cell signalling; insecticide;  
XX  
KW pharmaceutical; gene; ds.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX

PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
XX Claim 1; SEQ ID NO 733; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signaling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 5356 BP; 1275 A; 1352 C; 1364 G; 1365 T; 0 U; 0 Other;  
SQ  
XX  
XX Query Match 84.4%; Score 15.2; DB 4; Length 5356;  
XX Best Local Similarity 87.5%; Pred. No. 1e+03;  
XX Matches 14; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 2 ACCACGTSCCGGCAT 17  
DB 377 ACCACGTSCCGGCAT 362  
XX  
XX RESULT 25  
XX ABD32859/C  
XX ID ABD32859 standard; DNA; 110021 BP.  
XX AC ABD32859;  
XX  
XX 18-NOV-2004 (first entry)  
XX  
XX Mouse cancer-associated genomic DNA MD17-042.  
XX  
XX Mouse; ds; cancer-associated protein; gene; cytostatic; cancer;  
XX leukaemia; lymphoma; CAP.  
XX  
XX Mus musculus.  
XX  
XX WO2004074320-A2.  
XX  
XX PD 02-SEP-2004.  
XX  
XX 17-FEB-2004; 2004WO-US004730.  
XX  
XX 14-FEB-2003; 2003US-00367094.  
XX 14-MAR-2003; 2003US-00388836.  
XX 15-APR-2003; 2003US-00417375.  
XX 13-JUN-2003; 2003US-00461862.  
XX 15-SEP-2003; 2003US-00663431.  
XX 15-DEC-2003; 2003US-00737318.  
XX  
XX (SAGR-) SAGRES DISCOVERY INC.  
XX  
XX Morris DW, Morris DW, Malandro MS;  
XX  
XX WPI; 2004-652914/63.  
XX  
XX New isolated cancer-associated polynucleotides and polypeptides useful  
PT for diagnosing, preventing or treating cancers, especially lymphoma and

PT leukemia, or in screening for agents that modulate cancer.  
XX  
XX disclosure; seqid 533; 310pp; English.  
PS  
XX  
XX The invention relates to an isolated nucleic acid comprising at least 10  
CC contiguous nucleotides of any of the 233 polynucleotide sequences given  
CC in the specification, or its complement. The nucleic acids encode cancer-  
CC associated proteins. Also included are an expression vector comprising  
CC the isolated nucleic acid cited above, a host cell comprising the above  
CC recombinant nucleic acid or expression vector, a microarray for detecting  
CC a cancer-associated (CA) nucleic acid comprising at least one probe  
CC comprising at least 10 contiguous nucleotides of any of the above-  
CC mentioned nucleotide sequences, an isolated polypeptide (encoded within  
CC an open reading frame of a CA sequence selected from any of the 95  
CC polynucleotide sequences as mentioned in the specification, or its  
CC complement), an isolated antibody, (or its antigen binding fragment) that  
CC binds to the above polypeptide, a hybridoma that produces the above  
CC monoclonal antibody, a pharmaceutical composition comprising the above  
CC antibody and a pharmaceutical excipient, a kit for detecting cancer  
CC cells (comprising the antibody cited above, methods for diagnosing cancer  
CC or for detecting the presence or absence of cancer cells in an  
CC individual, a method for inhibiting growth of cancer cells in an  
CC individual, a method for delivering a therapeutic agent to cancer cells  
CC in an individual, an electronic library comprising the above  
CC polynucleotide or polypeptide (or their fragments), methods of screening  
CC for anticancer activity or for a bioactive agent capable of modulating  
CC the activity of a CA protein (CAP), methods for detecting cancer  
CC associated with expression of a polypeptide in a test cell sample, a  
CC method for treating cancers and a method for inhibiting the expression of  
CC CA gene in a cell. The composition and methods are useful for detecting,  
CC diagnosing, preventing and treating cancer, especially lymphoma and  
CC leukaemia. These may also be used in screening for agents that modulate  
CC cancer. The present sequence is a mouse CAP genomic sequence. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 110021 BP; 27669 A; 24618 C; 25713 G; 30636 T; 0 U; 1385 Other;  
SQ  
XX  
XX Query Match 84.4%; Score 15.2; DB 13; Length 110021;  
XX Best Local Similarity 87.5%; Pred. No. 1e+03;  
XX Matches 14; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 3 CCACGTSCCGGCATG 18  
DB 41825 CCACGTSCCGGCATG 41810  
XX  
XX RESULT 26  
XX ABL75000  
XX ID ABL75000 standard; cDNA; 209 BP.  
XX AC ABL75000;  
XX  
XX 14-MAY-2002 (first entry)  
XX  
XX Corn tassal-derived polynucleotide (cdps) SEQ ID NO:4374.  
XX  
XX Corn; corn tassal-derived polynucleotide; cdps; hybrid breeding; CDPS;  
XX inheritance; characteristic; growth; development; disease resistance;  
XX environmental adaptability; quality; yield; molecular marker;  
XX multigene trait; plant breeding; corn tassal; gene; ss.  
XX  
XX Zea mays.  
XX  
XX US2001051335-A1.  
XX  
XX 13-DEC-2001.  
XX  
XX 16-APR-1999; 99US-00294093.  
XX  
XX 21-APR-1998; 98US-0082567P.  
XX



PA (LALG/) LALGUDI R. V.  
 PA (ITOL/) ITO L. Y.  
 PA (SHER/) SHERMAN B. K.  
 PI Laigudi RV, Ito LY, Sherman BK;  
 DR WPI; 2002-163647/21.  
 XX  
 XX Novel purified corn taase1-derived polynucleotide useful for determining  
 PT altered gene expression, to recover regulatory elements and to follow  
 PT inheritance of desirable characteristics through hybrid breeding  
 PT programs.  
 PS  
 PS Claim 1; SEQ ID NO 4374; 201pp; English.  
 XX  
 XX The present sequence describes a purified corn taase1-derived  
 CC polynucleotide sequence (cdps) comprising a nucleic acid sequence  
 CC selected from those given in ABL70627 to ABL76833. The cdps sequences  
 CC encode corn taase1-derived polypeptides (CDPs). The cdps sequences (1)  
 CC can be used for determining altered gene expression, to recover  
 CC regulatory elements and to follow inheritance of desirable  
 CC characteristics through hybrid breeding programs. (1) are also useful in  
 CC the evaluation and alteration of desired characteristics associated with  
 CC growth and development, disease resistance, environmental adaptability,  
 CC quality and yield, and as molecular markers for studying inheritance of  
 CC multigene traits in a plant breeding program. (1) can be used to produce  
 CC a taase1-specific profile of gene transcription, a transcript image, to  
 CC clone regulatory elements for use in transformation vectors, to express a  
 CC polypeptide, to identify, isolate or extend identical or related corn  
 CC taase1 nucleic acid sequences from DNA libraries, in nucleic acid  
 CC hybridisation or amplification technologies, as query sequences to  
 CC determine homology of known sequences, as probe for use in Southern or  
 CC Northern hybridisation, and to identify the presence of and/or to  
 CC determine the degree of similarity between two (or more) nucleic acid  
 CC sequences  
 XX  
 SQ Sequence 209 BP; 52 A; 62 C; 49 G; 42 T; 0 U; 4 Other;  
 Query Match 81.1%; Score 14.6; DB 6; Length 209;  
 Best Local Similarity 82.4%; Pred. No. 2.4e+03;  
 Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GACCACTGTCGCGCAT 17  
 Db 87 GACTACGTCGCGCAT 103  
 RESULT 27  
 ADQ52850  
 ID ADQ52850 standard; DNA; 520 BP.  
 XX  
 AC ADQ52850;  
 XX  
 DT 21-OCT-2004 (first entry)  
 DE Novel canine microarray-related DNA sequence SeqID4152.  
 XX  
 XX canine microarray; drug screening; toxicity assay;  
 KM environmental pollutant; cellular response; gene expression profile;  
 KM toxic response; liver necrosis; fatty liver disease;  
 KM protein adduct formation; hepatitis; dog; ds.  
 XX  
 OS Canis familiaris.  
 PN WO2004063324-A2.  
 XX  
 PD 29-JUN-2004.  
 XX  
 PF 05-MAY-2003; 2003WO-US013853.  
 XX  
 PR 03-MAY-2002; 2002US-0377240P.  
 XX  
 PA (GENE-) GENE LOGIC INC.

PA (PRIZ ) PFIZER PROD INC.  
 XX  
 XX Diggs JC, Porter M, Wei T;  
 PI  
 XX WPI; 2004-561890/54.  
 DR  
 XX  
 XX New isolated nucleic acid molecule, useful for drug screening and  
 PT toxicity assays or for assessing the impact, including toxicity, of a  
 PT compound, pharmaceutical agent or environmental pollutant on a cell or  
 PT living organism.  
 PS  
 PS Claim 1; SEQ ID NO 4152; 41pp; English.  
 XX  
 XX This invention is related to a novel isolated canine nucleic acid  
 CC sequences and the construction of canine microarrays containing a  
 CC significant portion of the canine genome. The isolated canine nucleic  
 CC acid sequences of the invention may be useful for drug screening and  
 CC toxicity assays. The invention is therefore useful for assessing the  
 CC impact, including toxicity, of a compound, pharmaceutical agent or  
 CC environmental pollutant on a cell or living organism. The methods are  
 CC useful for detecting genes that are up- or down-regulated in canines in a  
 CC disease state. The sequences are useful as diagnostic agents or markers  
 CC to detect a cellular response in a sample individually or as part of a  
 CC gene expression profile. It is also useful as a target for agents that  
 CC modulate gene expression or activity. The database is useful for  
 CC producing electronic Northern blots that allow the user to determine the cell  
 CC type or tissue in which a given gene is expressed and to allow  
 CC determination of the abundance or expression level of a given gene in a  
 CC particular tissue or cell. The methods are useful for determining the  
 CC similarity of a toxic response to one or more individual compounds. The  
 CC methods are useful for predicting at least one toxic response or the  
 CC likelihood that a compound or test agent will induce various specific  
 CC pathologies such as those of the liver (liver necrosis, fatty liver  
 CC disease, protein adduct formation or hepatitis), those of the kidney,  
 CC heart, brain or testes, or other pathologies associated with at least one  
 CC of the toxins. The methods are also useful for predicting or elucidating  
 CC the potential cellular pathways influenced, induced or modulated by the  
 CC compound or test agent due to the similarity of the expression profile  
 CC compared to the profile induced by a known toxin. The present sequence is  
 CC that of a canine DNA sequence which was claimed for use during the  
 CC production of a canine microarray of the invention.  
 XX  
 SQ Sequence 520 BP; 72 A; 197 C; 156 G; 76 T; 0 U; 19 Other;  
 Query Match 81.1%; Score 14.6; DB 13; Length 520;  
 Best Local Similarity 82.4%; Pred. No. 2.4e+03;  
 Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GACCACTGTCGCGCAT 17  
 Db 162 GGCACGTCGCGCAT 178  
 RESULT 28  
 AAC56488/c  
 ID AAC56488 standard; DNA; 521 BP.  
 XX  
 AC AAC56488;  
 XX  
 DT 25-JAN-2001 (first entry)  
 DE Eucalyptus grandis transcription factor DNA sequence #359.  
 XX  
 XX plant; transcription factor; gene expression; eucalyptus; pine; acacia;  
 KM poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;  
 KM basic helix-loop-helix zipper; homeotic; homeodomain; MADS;  
 KM homeodomain zipper; LIM domain; AP2, EREBs; zinc finger domain;  
 KM type 2 Cys2His2; CCAAT box element; MYB; ss.  
 XX  
 OS Eucalyptus grandis.  
 PN WO200053724-A2.

PD 14-SEP-2000.  
XX  
XX 09-MAR-2000; 2000WO-US006112.  
XX  
XX 11-MAR-1999; 99US-00266513.  
XX 18-AUG-1999; 99US-0149485P.  
XX  
PA (GENE-) GENESIS RES & DEV CORP LTD.  
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.  
XX  
XX Wood M, McGrath A, Shenk MA, Glenn M;  
XX WPI; 2000-579369/54.  
XX  
XX New isolated polynucleotide encoding a plant transcription factor for  
XX producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
XX having modified gene expression or modified activity of a polypeptide.  
XX  
XX Claim 1; Page 447; 747pp; English.  
XX  
XX The present invention relates to novel plant transcription factors from  
XX Eucalyptus grandis or Pinus radiata. The present sequence is the coding  
XX sequence for one such transcription factor. The transcription factor may  
XX be used to produce a plant having modified gene expression such as a  
XX woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or  
XX mahogany species or to modify the activity of a polypeptide in a plant.  
XX The transcription factors of the present invention are members from the  
XX following families of regulatory proteins: bZIP, bZIP family of G-box  
XX binding factors, basic helix-loop-helix zipper,  
XX homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2  
XX and ERBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements and  
XX MYB  
SQ Sequence 521 BP; 99 A; 197 C; 141 G; 84 T; 0 U; 0 Other;  
Query Match 81.1%; Score 14.6; DB 3; Length 521;  
Best Local Similarity 82.4%; Pred. No. 2.4e+03;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GACCACTGTCGCGCAT 17  
DB 497 GACCACTGTCGCGCAT 481  
RESULT 29  
AAC56107/c  
ID AAC56107 standard; DNA; 521 BP.  
XX  
XX AAC56107;  
AC  
XX  
XX 25-JAN-2001 (first entry)  
DT  
XX  
XX Eucalyptus grandis transcription factor DNA sequence #238.  
DE  
XX  
XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;  
XX poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;  
XX basic helix-loop-helix zipper; homeotic; homeodomain; homeobox;  
XX homeodomain zipper; LIM domain; AP2; ERBs; zinc finger domain;  
XX type 2 Cys2His2; CCAAT box element; MYB; ss.  
XX  
XX Eucalyptus grandis.  
OS  
XX  
XX WO200053724-A2.  
PN  
XX  
XX 14-SEP-2000.  
PD  
XX  
XX 09-MAR-2000; 2000WO-US006112.  
PF  
XX  
XX 11-MAR-1999; 99US-00266513.  
PR  
XX 18-AUG-1999; 99US-0149485P.  
XX  
XX (GENE-) GENESIS RES & DEV CORP LTD.  
PA  
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX  
XX Wood M, McGrath A, Shenk MA, Glenn M;  
XX WPI; 2000-579369/54.  
XX  
XX New isolated polynucleotide encoding a plant transcription factor for  
XX producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
XX having modified gene expression or modified activity of a polypeptide.  
XX  
XX Claim 1; Page 109; 747pp; English.  
XX  
XX The present invention relates to novel plant transcription factors from  
XX Eucalyptus grandis or Pinus radiata. The present sequence is the coding  
XX sequence for one such transcription factor. The transcription factor may  
XX be used to produce a plant having modified gene expression such as a  
XX woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or  
XX mahogany species or to modify the activity of a polypeptide in a plant.  
XX The transcription factors of the present invention are members from the  
XX following families of regulatory proteins: bZIP, bZIP family of G-box  
XX binding factors, basic helix-loop-helix zipper,  
XX homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2  
XX and ERBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements and  
XX MYB  
SQ Sequence 521 BP; 99 A; 197 C; 141 G; 84 T; 0 U; 0 Other;  
Query Match 81.1%; Score 14.6; DB 3; Length 521;  
Best Local Similarity 82.4%; Pred. No. 2.4e+03;  
Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
OY 1 GACCACTGTCGCGCAT 17  
DB 497 GACCACTGTCGCGCAT 481  
RESULT 30  
ADA48826  
ID ADA48826 standard; DNA; 566 BP.  
XX  
XX ADA48826;  
AC  
XX  
XX 20-NOV-2003 (first entry)  
DT  
XX  
XX Banana gene conferring disease resistance in plants.  
DE  
XX  
XX disease resistance; pathogen tolerance; plant pathogen; ds; gene; plant;  
XX banana.  
XX  
XX Musa sp.  
OS  
XX  
XX WO200300906-A2.  
PN  
XX  
XX 03-JAN-2003.  
PD  
XX  
XX 21-JUN-2002; 2002WO-1B002453.  
PF  
XX  
XX 22-JUN-2001; 2001US-0300112P.  
PR 26-SEP-2001; 2001US-0352277P.  
XX 22-MAR-2002; 2002US-0366535P.  
XX  
XX (SYGN) SYNGENTA PARTICIPATIONS AG.  
PA  
XX  
XX Glazebrook J, Briggs S, Cooper B, Goff SA, Moughamer T;  
XX Katagiri F, Kreps J, Provart N, Rieke D, Zhu T;  
XX WPI; 2003-184052/18.  
XX  
XX New polynucleotide comprising a plant nucleotide sequence having an open  
XX reading frame that encodes a polypeptide associated with disease  
XX resistance, useful for conferring resistance or tolerance to a plant  
XX pathogen.  
XX  
XX Disclosure; SEQ ID NO 896; 299pp; English.  
PS

```

XX  The invention relates to a novel isolated polynucleotide comprising a
CC  plant nucleotide sequence having an open reading frame that encodes a
CC  polypeptide associated with disease resistance or its fragment having
CC  substantially the same activity as the full-length polypeptide. The
CC  polynucleotide of the invention is useful for conferring resistance or
CC  tolerance to a plant pathogen. The present sequence represents a gene
CC  conferring disease resistance used in the invention.
XX
SQ  Sequence 568 BP; 145 A; 154 C; 136 G; 128 T; 0 U; 5 Other;
    Query March      81.1%; Score 14.6; DB 9; Length 568;
    Best Local Similarity 82.4%; Pred. No. 2.4e+03;
    Matches 14; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY  1 GACCACGTCGCGCAT 17
    |||||:|||||
Db  378 GACCACGTCGCGCAT 394

Search completed: July 20, 2005, 16:11:31
Job time : 331.143 secs

```

**THIS PAGE IS BLANK**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 20, 2005, 14:42:10 ; Search time 852.429 Seconds

(without alignments)  
1023.187 Million cell updates/sec

Title: US-10-017-471b-1

Perfect score: 18

Sequence: 1 gaccacgtaccsagcagc 18

Scoring table: IDENTITY NUC

Gap: 10.0, Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 300 summaries

Database : GenEmbl.\*  
1: gb\_da.\*  
2: gb\_hcg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_scs.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17.2	95.6	1203	1	STMAFSA
2	17.2	95.6	12070	1	AB011413
3	17.2	95.6	155628	1	AB016876
4	17.2	95.6	296300	1	AB005035
5	16	88.9	1215	1	AB001608
6	16	88.9	2234	1	AB001683
7	16	88.9	6727	1	SC0007731
8	16	88.9	290850	1	SC039127
9	15.6	86.7	513	6	BD139044
10	15.6	86.7	513	6	AR198867
11	15.6	86.7	744	6	CQ585804
12	15.6	86.7	762	6	BD139042
13	15.6	86.7	762	6	AR198865
14	15.6	86.7	859	3	AY060758
15	15.6	86.7	1008	6	AR480461
16	15.6	86.7	1188	6	AX567274
17	15.6	86.7	1408	6	E30250
18	15.6	86.7	1408	6	E30254
19	15.6	86.7	1408	6	AR215256

20	15.6	86.7	1408	6	AR215257
21	15.6	86.7	1798	8	NC425011
22	15.6	86.7	2000	6	AX656460
23	15.6	86.7	2048	6	BD139028
24	15.6	86.7	2048	6	AR198851
25	15.6	86.7	2395	6	CO789114
26	15.6	86.7	2395	6	AR218973
27	15.6	86.7	2395	6	BD003885
28	15.6	86.7	3233	9	HS112B
29	15.6	86.7	4927	9	AF227905
30	15.6	86.7	5042	6	AX405692
31	15.6	86.7	6252	1	SVU24659
32	15.6	86.7	6684	9	BC041098
33	15.6	86.7	10127	1	AE011135
34	15.6	86.7	10643	1	AE005070
35	15.6	86.7	11333	1	AE008410
36	15.6	86.7	11790	1	AE011835
37	15.6	86.7	12332	6	CO585803
38	15.6	86.7	12606	1	AE005995
39	15.6	86.7	13121	1	AE004608
40	15.6	86.7	14385	1	AE007344
41	15.6	86.7	45752	3	DMBN3381
42	15.6	86.7	72284	2	AC020328
43	15.6	86.7	77428	3	AC017634
44	15.6	86.7	80389	3	AC005643
45	15.6	86.7	103124	9	AL139818
46	15.6	86.7	103714	2	AC149664
47	15.6	86.7	110000	1	AP006618
48	15.6	86.7	112911	9	AC005865
49	15.6	86.7	115932	9	AC011446
50	15.6	86.7	124950	8	AC108059
51	15.6	86.7	131246	8	OSJND029
52	15.6	86.7	150089	3	AC144743
53	15.6	86.7	154681	3	AC099008
54	15.6	86.7	156573	2	AC073075
55	15.6	86.7	161569	3	AC104145
56	15.6	86.7	163424	2	AC129293
57	15.6	86.7	165696	9	AL1353803
58	15.6	86.7	168519	8	AC144737
59	15.6	86.7	170929	10	AC130548
60	15.6	86.7	171222	2	AC091387
61	15.6	86.7	175936	2	SPNBT1908
62	15.6	86.7	176383	8	OSJND003
63	15.6	86.7	178062	9	CNS01DRK
64	15.6	86.7	186607	2	AC140777
65	15.6	86.7	190177	2	AC009959
66	15.6	86.7	210614	1	AB088224
67	15.6	86.7	210973	2	AC130949
68	15.6	86.7	211397	2	AC115177
69	15.6	86.7	215050	10	AL645854
70	15.6	86.7	229050	1	AL646057
71	15.6	86.7	229768	2	AC095653
72	15.6	86.7	229635	2	AC094256
73	15.6	86.7	232042	2	AC128144
74	15.6	86.7	240904	2	AC126204
75	15.6	86.7	242135	2	AC127842
76	15.6	86.7	247254	1	AC114844
77	15.6	86.7	260050	1	SME591782
78	15.6	86.7	265472	3	AE003817
79	15.6	86.7	269267	1	AC110861
80	15.6	86.7	299050	1	SC039104
81	15.6	86.7	309052	3	AE003425
82	15.6	86.7	347786	1	BX640438
83	15.6	86.7	349146	1	BX640424
84	15.6	86.7	349980	6	AX571760
85	15.6	86.7	39284	1	AB117719
86	15.6	86.7	4356	6	CO594113
87	15.2	84.4	94874	10	AL808126
88	15.2	84.4	109151	6	AC017637
89	15.2	84.4	110021	6	CO870112
90	15.2	84.4	161278	3	AC008223
91	15.2	84.4	181021	10	AL663031
92	15.2	84.4	184889	9	AL442125

AR215257	Sequence
AR250111	Nitella c
AX656460	Sequence
BD139028	Virulence
AR198851	Sequence
CO789114	Sequence
AR218973	Sequence
BD003885	Polynucle
X81346	Homo sapien
AF227905	Homo sapi
AX405692	Sequence
U24659	Streptomyce
BC041098	Homo sapi
AE011135	Metanorosa
AE005070	Halobacte
AE008410	Streptoco
AE011835	Xanthomon
CO585803	Sequence
AE005995	Callobact
AE004608	Pseudomon
AE007344	Streptoco
AL133506	Drosophil
AC020328	Drosophil
AC017634	Drosophil
AC005643	Drosophil
AL139818	Human DNA
AC149664	Caroli
AC005865	Continuatio
AC011446	Homo sapi
AC108059	Homo sapi
BX842605	Oryza sat
AC144743	Oryza sat
AC099008	Drosophil
AC073075	Homo sapi
AC104145	Drosophil
AC129293	Mus muscu
AL1353803	Human DNA
AC144737	Oryza sat
AC130548	Mus muscu
AC091387	Drosophil
AL44930	Streptoco
AL606460	Oryza sat
AL117356	Human chr
AC140777	Mus muscu
AC009959	Homo sapi
AB088224	Streptomy
AC130949	Rattus no
AC115177	Mus muscu
AL645854	Mus muscu
AL646057	Rattus no
AC095653	Rattus no
AC094256	Rattus no
AC128144	Rattus no
AC126204	Rattus no
AC127842	Rattus no
AL591782	Sinorhizo
AE003817	Drosophil
AC110861	Rattus no
AL939104	Streptomy
AE003425	Drosophil
BX640438	Bordetell
BX640424	Bordetell
AX571760	Sequence
AB117719	Rhodococc
CO594113	Sequence
AL808126	Mouse DNA
AC017637	Drosophil
CO870112	Sequence
AC008223	Drosophil
AL663031	Mouse DNA
AL442125	Human DNA

C 93	15.2	84.4	185639	2	AC145465	AC145465	Callipectu	C 166	14.6	81.1	2567	10	AK098097	AK098097	Mus muscu
C 94	15.2	84.4	190153	3	AC008201	AC008201	Drosophill	C 167	14.6	81.1	2761	9	BC043916	BC043916	Mus muscu
C 95	15.2	84.4	206880	2	AC128103	AC128103	Rattus no	C 168	14.6	81.1	2879	6	C0722672	C0722672	Sequence
C 96	15.2	84.4	239972	2	AC095835	AC095835	Rattus no	C 169	14.6	81.1	2883	1	SY0ATBP	SY0ATBP	Synecoccc
C 97	15.2	84.4	244416	3	AE003744	AE003744	Drosophill	C 170	14.6	81.1	2995	6	C0580250	C0580250	Sequence
C 98	15.2	84.4	250902	2	AC131965	AC131965	Rattus no	C 171	14.6	81.1	3290	3	DME011320	DME011320	Drosophill
C 99	15.2	84.4	292200	1	SC0939129	SC0939129	Streptomy	C 172	14.6	81.1	3375	8	AF169793	AF169793	Podospira
C 100	15.2	84.4	299225	1	AP005045	AP005045	Streptomy	C 173	14.6	81.1	3380	6	C0580244	C0580244	Sequence
C 101	15.2	84.4	300140	1	AP005556	AP005556	Bradyrhiz	C 174	14.6	81.1	3500	6	C0572837	C0572837	Sequence
C 102	15.2	84.4	300297	1	AE016944	AE016944	Bacteroid	C 175	14.6	81.1	3542	1	AF033105	AF033105	Sequence
C 103	15	83.3	309	13	AF427374	AF427374	Unculture	C 176	14.6	81.1	3630	3	HSMB04113	HSMB04113	Sequence
C 104	15	83.3	309	13	AF427376	AF427376	Unculture	C 177	14.6	81.1	4054	3	MSU05719	MSU05719	Homo sapi
C 105	15	83.3	309	13	AF427378	AF427378	Unculture	C 178	14.6	81.1	4360	6	C0602123	C0602123	Sequence
C 106	15	83.3	576	11	BV030613	BV030613	Unculture	C 179	14.6	81.1	4421	9	HUMCATP	HUMCATP	Human plasm
C 107	15	83.3	39737	9	AC011521	AC011521	Homo sapi	C 180	14.6	81.1	4789	9	AB046776	AB046776	Homo sapi
C 108	15	83.3	48109	2	AC151741	AC151741	Rattus no	C 181	14.6	81.1	4979	9	HUMATP282X	HUMATP282X	Human plasm
C 109	15	83.3	66256	2	AC087454	AC087454	Homo sapi	C 182	14.6	81.1	5137	14	AY044867	AY044867	Hepatitis
C 110	15	83.3	72229	2	AC101639	AC101639	Mus muscu	C 183	14.6	81.1	6037	10	AK122256	AK122256	Mus muscu
C 111	15	83.3	135975	9	AC105935	AC105935	Mus muscu	C 184	14.6	81.1	6676	1	AF031590	AF031590	Sequence
C 112	15	83.3	135439	9	AC008626	AC008626	Homo sapi	C 185	14.6	81.1	6755	1	AF002247	AF002247	Rhodococc
C 113	15	83.3	163424	2	AC099330	AC099330	Homo sapi	C 186	14.6	81.1	6946	6	AX818197	AX818197	Sequence
C 114	15	83.3	167471	9	AC019211	AC019211	Homo sapi	C 187	14.6	81.1	6946	9	AB007919	AB007919	Homo sapi
C 115	15	83.3	167722	9	AC073548	AC073548	Homo sapi	C 188	14.6	81.1	7838	6	C0601595	C0601595	Sequence
C 116	15	83.3	174924	9	AC022145	AC022145	Homo sapi	C 189	14.6	81.1	8245	1	AE014770	AE014770	Sequence
C 117	15	83.3	187723	9	AC016889	AC016889	Homo sapi	C 190	14.6	81.1	8324	2	AC019926	AC019926	Drosophill
C 118	15	83.3	189149	2	AC046171	AC046171	Homo sapi	C 191	14.6	81.1	8931	1	AF161183	AF161183	Burkholde
C 119	15	83.3	204746	2	AC055866	AC055866	Homo sapi	C 192	14.6	81.1	9198	1	AB002355	AB002355	Homo sapi
C 120	15	83.3	217325	10	AC115360	AC115360	Mus muscu	C 193	14.6	81.1	10023	1	AE009262	AE009262	Agrobacte
C 121	15	83.3	239718	2	AC101774	AC101774	Mus muscu	C 194	14.6	81.1	10235	1	AE008330	AE008330	Agrobacte
C 122	15	83.3	252486	2	AC111445	AC111445	Rattus no	C 195	14.6	81.1	10769	1	AE000881	AE000881	Methanoba
C 123	15	83.3	253946	2	AC105161	AC105161	Mus muscu	C 196	14.6	81.1	10970	1	AE001869	AE001869	Deinococc
C 124	15	83.3	254939	2	AC099076	AC099076	Rattus no	C 197	14.6	81.1	11188	6	AX339644	AX339644	Sequence
C 125	15	83.3	273325	4	AF324420	AF324420	Equus cab	C 198	14.6	81.1	11188	6	AX275002	AX275002	Sequence
C 126	14.6	81.1	465	9	HS338572	HS338572	Homo sapi	C 199	14.6	81.1	12198	6	AF187159	AF187159	Streptomy
C 127	14.6	81.1	510	6	AX425418	AX425418	Sequence	C 200	14.6	81.1	12404	1	AE014736	AE014736	Bitfidbac
C 128	14.6	81.1	510	6	AX986112	AX986112	Sequence	C 201	14.6	81.1	13221	1	AE014422	AE014422	Brucella
C 129	14.6	81.1	510	6	BD120971	BD120971	EST and e	C 202	14.6	81.1	13453	1	AE009518	AE009518	Sequence
C 130	14.6	81.1	510	6	BD120971	BD120971	EST and e	C 203	14.6	81.1	13537	9	AF257737	AF257737	Homo sapi
C 131	14.6	81.1	568	6	AX660539	AX660539	Sequence	C 204	14.6	81.1	13816	9	HS404468	HS404468	Homo sapi
C 132	14.6	81.1	635	5	AY394439	AY394439	Spheroelid	C 205	14.6	81.1	14069	2	AC019558	AC019558	Drosophill
C 133	14.6	81.1	766	8	AF051369	AF051369	Oryza sat	C 206	14.6	81.1	14704	1	AE001970	AE001970	Deinococc
C 134	14.6	81.1	808	11	G06627	G06627	human STS W	C 207	14.6	81.1	15836	1	AE005057	AE005057	Halobacte
C 135	14.6	81.1	808	11	G10673	G10673	human STS C	C 208	14.6	81.1	17083	1	SC1536588	SC1536588	Streptomy
C 136	14.6	81.1	888	6	AR123908	AR123908	Sequence	C 209	14.6	81.1	17083	6	AX593213	AX593213	Sequence
C 137	14.6	81.1	913	8	AK119706	AK119706	Oryza sat	C 210	14.6	81.1	18524	6	CQ730656	CQ730656	Sequence
C 138	14.6	81.1	915	8	AK064673	AK064673	Oryza sat	C 211	14.6	81.1	20435	9	HS402535	HS402535	Homo sapi
C 139	14.6	81.1	921	6	C0580251	C0580251	Sequence	C 212	14.6	81.1	21051	2	AC017802	AC017802	Drosophill
C 140	14.6	81.1	922	3	AY051697	AY051697	Drosophill	C 213	14.6	81.1	28222	2	AC019963	AC019963	Drosophill
C 141	14.6	81.1	1004	10	AF056244	AF056244	Mus muscu	C 214	14.6	81.1	28222	1	AY233211	AY233211	Streptomy
C 142	14.6	81.1	1005	6	CQ445545	CQ445545	Sequence	C 215	14.6	81.1	34667	9	HS4314905	HS4314905	Homo sapi
C 143	14.6	81.1	1041	6	AX573922	AX573922	Sequence	C 216	14.6	81.1	36615	1	AF449411	AF449411	Myxococcu
C 144	14.6	81.1	1068	6	AX553318	AX553318	Sequence	C 217	14.6	81.1	37108	3	CET20F10	CET20F10	Caenorhabdi
C 145	14.6	81.1	1173	10	NM2427265	NM2427265	Sequence	C 218	14.6	81.1	40505	6	AF574197	AF574197	Sequence
C 146	14.6	81.1	1215	3	AY122204	AY122204	Drosophill	C 219	14.6	81.1	49899	10	AC134735	AC134735	Rattus no
C 147	14.6	81.1	1242	3	U00790	U00790	Drosophill	C 220	14.6	81.1	50000	6	AX535264	AX535264	Sequence
C 148	14.6	81.1	1242	6	CQ572838	CQ572838	Sequence	C 221	14.6	81.1	50000	6	AX535266	AX535266	Sequence
C 149	14.6	81.1	1424	6	AR510109	AR510109	Sequence	C 222	14.6	81.1	59816	1	AF333038	AF333038	Streptomy
C 150	14.6	81.1	1435	8	AK105163	AK105163	Oryza sat	C 223	14.6	81.1	61125	2	AC137716	AC137716	Homo sapi
C 151	14.6	81.1	1630	5	BC072173	BC072173	Xenopus 1	C 224	14.6	81.1	61371	2	AC105135	AC105135	Drosophill
C 152	14.6	81.1	1649	10	BC012206	BC012206	Mus muscu	C 225	14.6	81.1	62001	2	AC015206	AC015206	Drosophill
C 153	14.6	81.1	1722	14	PYU19303	PYU19303	Prairie vol	C 226	14.6	81.1	64291	2	AC013683	AC013683	Homo sapi
C 154	14.6	81.1	1749	6	AR339656	AR339656	Sequence	C 227	14.6	81.1	82972	8	NCB24H17	NCB24H17	Neurospor
C 155	14.6	81.1	1749	6	AX275014	AX275014	Sequence	C 228	14.6	81.1	95579	9	AC023162	AC023162	Homo sapi
C 156	14.6	81.1	1793	10	AF031035	AF031035	Mus muscu	C 229	14.6	81.1	98356	9	HS434J2	HS434J2	Homo sapi
C 157	14.6	81.1	1898	1	SCTRXABGN	SCTRXABGN	Streptomyce	C 230	14.6	81.1	101777	9	AC105210	AC105210	Homo sapi
C 158	14.6	81.1	1898	6	BD179569	BD179569	Highly th	C 231	14.6	81.1	104947	9	AC008116	AC008116	Homo sapi
C 159	14.6	81.1	2397	1	MFO296160	MFO296160	Mycobacte	C 232	14.6	81.1	110000	1	AE000516	AE000516	Sequence
C 160	14.6	81.1	2403	10	DB6177	DB6177	Mouse mRNA	C 233	14.6	81.1	110000	1	AE016822	AE016822	Sequence
C 161	14.6	81.1	2519	6	CQ841741	CQ841741	Sequence	C 234	14.6	81.1	110000	1	AE017180	AE017180	Sequence
C 162	14.6	81.1	2519	6	AK123359	AK123359	Homo sapi	C 235	14.6	81.1	110000	1	AE017282	AE017282	Sequence
C 163	14.6	81.1	2534	9	AX834822	AX834822	Sequence	C 236	14.6	81.1	110000	1	AP006840	AP006840	Sequence
C 164	14.6	81.1	2534	9	AK097489	AK097489	Homo sapi	C 237	14.6	81.1	110000	1	AP006840_02	AP006840_02	Sequence
C 165	14.6	81.1	2534	10	BC031774	BC031774	Mus muscu	C 238	14.6	81.1	110000	1	AP006840_26	AP006840_26	Sequence

C 239	14.6	81.1	110000	2	AC096315_6	Continuation (7 of
C 240	14.6	81.1	110000	2	AC15182_1	Continuation (2 of
C 241	14.6	81.1	110000	2	AC151851_0	AC151851 Yakiyuga
C 242	14.6	81.1	110000	2	AC151851_1	Continuation (2 of
C 243	14.6	81.1	110000	2	AP006494_0	AP006494 Cyanidios
C 244	14.6	81.1	110000	2	AP006500_00	AP006500 Cyanidios
C 245	14.6	81.1	110000	8	CR382121_05	Continuation (6 of
C 246	14.6	81.1	110000	8	CR382128_15	Continuation (16 of
C 247	14.6	81.1	115968	9	AL139100	AL139100 Homo sapi
C 248	14.6	81.1	118699	9	AC133004	AC133004 Homo sapi
C 249	14.6	81.1	121849	8	AC092779	AC092779 Oryza sat
C 250	14.6	81.1	126784	2	AY263454	AY263454 Sus scrofa
C 251	14.6	81.1	128000	9	AC098690	AC098690 Homo sapi
C 252	14.6	81.1	128638	9	AC015968	AC015968 Homo sapi
C 253	14.6	81.1	135964	9	AL353593	AL353593 Homo sapi
C 254	14.6	81.1	137952	2	AC105141	AC105141 Homo sapi
C 255	14.6	81.1	142329	2	AC128499	AC128499 Rattus no
C 256	14.6	81.1	146947	2	AC012515	AC012515 Homo sapi
C 257	14.6	81.1	148675	10	AC131769	AC131769 Mus muscu
C 258	14.6	81.1	153648	3	AC008210	AC008210 Drosophi
C 259	14.6	81.1	154110	2	AC148737	AC148737 Ornithorh
C 260	14.6	81.1	154736	2	AL139246	AL139246 Homo sapi
C 261	14.6	81.1	158892	2	CNS01R14	AL162871 Homo sapi
C 262	14.6	81.1	159880	10	AC142258	AC142258 Mus muscu
C 263	14.6	81.1	161096	9	AC016257	AC016257 Homo sapi
C 264	14.6	81.1	161117	3	AC008311	AC008311 Drosophi
C 265	14.6	81.1	162829	2	CR352283	CR352283 Dario rer
C 266	14.6	81.1	163455	2	AC009427	AC009427 Homo sapi
C 267	14.6	81.1	163579	9	AC009363	AC009363 Homo sapi
C 268	14.6	81.1	164396	2	HSAC000380	AC000380 Homo sapi
C 269	14.6	81.1	164766	2	AC026657	AC026657 Homo sapi
C 270	14.6	81.1	164959	2	AC087823	AC087823 Homo sapi
C 271	14.6	81.1	166007	2	CNS01DRC	AL117186 Human chr
C 272	14.6	81.1	166176	2	AC104143	AC104143 Drosophi
C 273	14.6	81.1	166176	2	AC104143	AC104143 Drosophi
C 274	14.6	81.1	166450	2	AC078874	AC078874 Homo sapi
C 275	14.6	81.1	167601	2	AC120208	AC120208 Canis fam
C 276	14.6	81.1	168448	3	AC009346	AC009346 Drosophi
C 277	14.6	81.1	169168	5	BX546502	BX546502 Zebrafish
C 278	14.6	81.1	169189	2	AC125293	AC125293 Drosophi
C 279	14.6	81.1	171032	2	AC091761	AC091761 Felis cat
C 280	14.6	81.1	171421	8	AP004776	AP004776 Oryza sat
C 281	14.6	81.1	172560	2	AC091794	AC091794 Felis cat
C 282	14.6	81.1	173441	9	AP000787	AP000787 Homo sapi
C 283	14.6	81.1	173702	3	AC007593	AC007593 Drosophi
C 284	14.6	81.1	174234	10	AC117098	AC117098 Rattus no
C 285	14.6	81.1	174612	2	AC023889	AC023889 Homo sapi
C 286	14.6	81.1	174673	9	AC019187	AC019187 Homo sapi
C 287	14.6	81.1	174906	9	AC067930	AC067930 Homo sapi
C 288	14.6	81.1	175996	2	AC108480	AC108480 Drosophi
C 289	14.6	81.1	176205	2	AC150984	AC150984 Bos tauru
C 290	14.6	81.1	176918	2	AC109613	AC109613 Bos tauru
C 291	14.6	81.1	180727	2	AC021174	AC021174 Homo sapi
C 292	14.6	81.1	180783	9	AC079611	AC079611 Homo sapi
C 293	14.6	81.1	181371	2	AC018893	AC018893 Homo sapi
C 294	14.6	81.1	181515	2	AC119068	AC119068 Canis fam
C 295	14.6	81.1	181904	3	AC008211	AC008211 Drosophi
C 296	14.6	81.1	182785	9	AC145937	AC145937 Pan trogl
C 297	14.6	81.1	186431	10	AC113092	AC113092 Mus muscu
C 298	14.6	81.1	188549	3	AC007818	AC007818 Drosophi
C 299	14.6	81.1	189050	1	AL646069	AL646069 Ralstonia
C 300	14.6	81.1	189317	9	HSJ190720	AL050335 Human DNA

## ALIGNMENTS

RESULT 1  
STMAFSA  
LOCUS  
DEFINITION  
ACCESSION

STMAFSA  
S.grieseus afesa gene encoding a possible A-factor biosynthesis  
protein.  
M24250

1203 bp DNA linear BCT 26-APR-1993

VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT  
FEATURES  
source

M24250.1 GI:153148  
A-factor biosynthesis.  
Streptomyces grieseus  
Streptomyces grieseus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomyces; Streptomycetaceae; Streptomyces.  
1 (bases 1 to 1203)  
Horinouchi, S., Suzuki, H., Nishiyama, M. and Bappu, T.  
Nucleotide sequence and transcriptional analysis of the  
Streptomyces grieseus gene (afsa) responsible for A-factor  
biosynthesis  
J. Bacteriol. 171 (2), 1206-1210 (1989)  
89123125  
2492509  
Original source text: S.grieseus DNA.  
Location/Qualifiers  
1..1203  
/organism="Streptomyces grieseus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1911"  
204..1109  
/note="afsa protein"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="AA026693.1"  
/db\_xref="GI:153149"  
/translation="MDAEVHPVGIEMVHTRTPEDAFRNMVRLGRDFAVEAVLP  
HDHFPAPVDLDHDLVAEMROAMLAFFAGYGIPLGYHLLTLDVCHPEHIG  
VGGPEPIEGVGLCDLKNRAGAPAGRGVAVHGRDRLAATGVAATRFSTPRAYRM  
RGDPVPEIGISLPETAVPAPAPAGRAGVAVHGRDRLAATGVAATRFSTPRAYRM  
HYPMLELBARQACLVAGPAGVAVHGRDRLAATGVAATRFSTPRAYRM  
TVAVTGHQDEIVFTVLSGPRAGH"

ORIGIN

Query Match 95.6%; Score 17.2; DB 1; Length 1203;  
Best Local Similarity 88.9%; Pred. No. 1.4e+03;  
Matches 16; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Cy 1 GACCACGTCSCGCGCATG 18  
Db 855 GACCACGTCGCGGCATG 872

RESULT 2  
AB011413  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
COMMENT  
FEATURES  
source

AB011413 12070 bp DNA linear BCT 07-AUG-1998  
Streptomyces grieseus genes for Orf2, Orf3, Orf4, Orf5, AfesA, Orf8,  
partial and complete cds.  
AB011413  
AB011413.1 GI:3401946  
Orf8; AfesA; Orf5; Orf4; Orf3; Orf2.  
Streptomyces grieseus  
Streptomyces grieseus  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomyces; Streptomycetaceae; Streptomyces.  
1 (sites)  
Uneyama, T.  
Open reading frame encoded around afsa gene  
Unpublished  
2 (bases 1 to 12070)  
Uneyama, T.  
Direct Submission  
Submitted (23-FEB-1998) Takashi Uneyama, University of Tokyo,  
Department of Agriculture and Life Sciences, Yayoi 1-1-1,  
Bunkyo-Ku, Tokyo 113, Japan  
(E-mail:aa67103@hongo.ecc.u-tokyo.ac.jp, Tel:+81-3-3812-2111)  
Location/Qualifiers  
1..12070  
/organism="Streptomyces grieseus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1911"  
complement(2660..3511)  
/gene="orf2"

CDS  
complement(2660..>3511)  
/gene="orf2"  
/function="sensor histidine kinase"  
/codon\_start=1  
/transl\_table=1  
/product="Orf2"  
/protein\_id="BAA32130.1"  
/db\_xref="GI:3401947"  
/translation="VAAPVPAHPGPEPAAPAAARLTAAATLTAVTLITVVGACLSNVLR  
TDDRMDAEDYEMKRRASVAALVYEDDIIRSGAPAEHRSVLVITGAGALTALA  
ALTGHVLSGRSGPAMEALQOEERLADAAHELRTPVAVMGSVYAAAGAGGOL  
PRIRARADRMADVENILTRGLEGRTRSGRSLRDOLVEYCAELPEGGPLESR  
LEESVYEDADLVAVARNILDMNVRHGPSRGAAGRRTAGDVRVTVADGNPAV  
APARLPS"  
complement(3523..>4197)  
/function="two-component regulator protein"  
/codon\_start=1  
/transl\_table=1  
/product="Orf3"  
/protein\_id="BAA32131.1"  
/db\_xref="GI:3401948"  
/translation="MRVLVVEDDDDLRVDVIGAGLRDGGFAMDCAWDPEADVLHLTA  
YDCVVLADVPVSGDTALPLEGRBRAGVGPVLCALDLSLBERLGRSGADVLAKP  
PSMRVLYRVGSLSRASARLPSFLGCADVYMDVABRHRVRRGVLISLSPREYVLOQ  
LLVHRDLYVTRTGLLHCHCDEMADPVSNVDAVAGLRKLGSGELVHTVVGCGFLLS  
ADPGPS"  
complement(4333..>5250)  
/function="methyltransferase"  
/codon\_start=1  
/transl\_table=1  
/product="Orf4"  
/protein\_id="BAA32132.1"  
/db\_xref="GI:3401949"  
/translation="WATQALGALRLGVFDRIGTGLERAEALAGSLGTHPRATLLRL  
ALAGQLSEBPAGFRTTAAGNVARADPGTMAMFTDPVWLKGMDDLSVTR  
GETTFDVTGDFEGHLEHREHLESAENFASOGTRLTAETVPHHIDRGEPRLVDIG  
GGDGLIASIIRAHGPRGV.FDTAEGLAQAPRLARGLDGRVTLFNGDPPASAPAG  
GDLVLIANSIHMDMDVARGHLEHRIHVDVLPDQSLIVEPVLPATVPADRPNVYISDL  
MLVAVGGERERADDFALCTAGFACGA"  
complement(5450..6361)  
/function="regulator protein"  
/codon\_start=1  
/transl\_table=1  
/product="Orf5"  
/protein\_id="BAA32133.1"  
/db\_xref="GI:3401950"  
/translation="MSELPDQVTRLAVNDEGTFPSAATALRLTPASVORVALAQ  
RTGRVLMRTKVPRTSGEVVALAROLARLEHQAALGMSGPEPTLLPIAVNSDS  
LATNSWPCGACRSMDSATTPAGGPGYGRPAAGADGKDDVAGGDDGLFGPAG  
RMRYLPVAPSGPADWLRDGTALRELIIGEPVVCFRDLDLDAFVRRLGPARPS  
ARRHLVPTSEGFANAVASGMGVPEVQAPPLSDGRLVRLAPEPTVDPLVYQOMK  
LESPALAAVEVAEAAALDEAPPN"  
7486..8391  
/codon\_start=1  
/transl\_table=1  
/product="Alfa"  
/protein\_id="BAA32134.1"  
/db\_xref="GI:3401951"  
/translation="MDAEAEVHPVIGIEMVHRTREDAFPRNMVVLGRDRAVEAVLP  
HDHPPAPVGGDLDHPLVIAEAMROALAPLAGYGIPIYHILLTELIDYCHPHGL  
VGGEPTEIGLEVFCSDDLKRRAGLPQCGYGAVAHNGDLAATGVAATFTSPKAYRR  
KGDVVEGISLEPIAPVAPSGARVADVLISGTGRGVWELRVDTTHPTLLPQKPN  
HVPGLMLEAARQAALVAGPAGIVPARTRFHRYSBFGSPCWIGAVVQGADEDTV  
TVRVYGHODGETVFSLSGPRAHG"  
10558..11598  
/function="alcohol dehydrogenase"  
/codon\_start=1  
/transl\_table=1  
/product="Orf8"  
/protein\_id="BAA32135.1"  
/db\_xref="GI:3401952"  
/translation="MTVAAAYAPAPAKAPLERTTERRPVGEPTLLIDIKFAGICHSD  
IHQARDGMBEGIFPVPQGEHLAGIYAEITGPVTRKVDKRGVGVGMVDSCGTCDACLM

CDS  
complement(155628  
/organism="Pseudomonas syringae pv. tomato str. DC3000"  
/mol\_type="genomic DNA"  
/strain="DC3000"  
/db\_xref="taxon:223283"  
/note="pathovar: tomato"  
complement(159..4811)  
/locus\_tag="PSPT05481"  
complement(159..4811)  
/locus\_tag="PSPT05481"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAO58900.1"  
/db\_xref="GI:28855842"  
/translation="MTSTPLQSNVHTAVAPAPASQQRSDLGIFCOALVGLTPVGEYH  
DKLANSILNOOLAPEROIAIHQLAIPVSHWDSNIHPTDPTAMALAGAEYLRM  
AIIHVPDGAWNTAIEDATVAPAGMALSAISIGEHRSLESILAAIYLMVY  
LHSAVALDESLPNKSLIDAPRDPDSTVYRGLIQRGMGLCPESGSLFRS  
PHATVHALCCTPAQFKSVFKPLSLTMDWFGSGSGHTSPVLTQALAGVYADHVAAYQ

CDS  
complement(155628  
/organism="Pseudomonas syringae pv. tomato str. DC3000"  
/mol\_type="genomic DNA"  
/strain="DC3000"  
/db\_xref="taxon:223283"  
/note="pathovar: tomato"  
complement(159..4811)  
/locus\_tag="PSPT05481"  
complement(159..4811)  
/locus\_tag="PSPT05481"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAO58900.1"  
/db\_xref="GI:28855842"  
/translation="MTSTPLQSNVHTAVAPAPASQQRSDLGIFCOALVGLTPVGEYH  
DKLANSILNOOLAPEROIAIHQLAIPVSHWDSNIHPTDPTAMALAGAEYLRM  
AIIHVPDGAWNTAIEDATVAPAGMALSAISIGEHRSLESILAAIYLMVY  
LHSAVALDESLPNKSLIDAPRDPDSTVYRGLIQRGMGLCPESGSLFRS  
PHATVHALCCTPAQFKSVFKPLSLTMDWFGSGSGHTSPVLTQALAGVYADHVAAYQ

ORIGIN		Query Match	95.6%; Score 17.2; DB 1; Length 12070;
		Best Local Similarity	88.9%; Pred. No. 9e+02;
		Matches	16; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Oy	1	GACCACGTCSCCGGCGATG 18	: : : : : :
Db	8137	GACCACGTCSCCGGCGATG 8154	: : : : : :
RESULT 3			
AE016876			
LOCUS			
DEFINITION			
Pseudomonas syringae pv. tomato str. DC3000 section 21 of 21 of the complete genome.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
1 (bases 1 to 155628)			
Buell, C.R., Joardar, V., Lindberg, M., Selengut, J., Paulsen, I.T., Gwyn, M.L., Dodson, R.J., Deboy, R.T., Durkin, A.S., Kolonay, J.F., Madupu, R., Daugherty, S., Brinkac, L., Beanan, M.J., Haft, D.H., Nelson, W.C., Davidson, T., Zafar, N., Zhou, L., Liu, J., Yuan, Q., Khouri, H., Fedorova, N., Tran, B., Russell, D., Berry, K., Uterback, T., Van Aken, S.E., Feldblyum, T.V., D'Ancenzo, M., Deng, W.L., Ramoa, A.R., Alfano, J.R., Catlinhour, S., Chatterjee, A.K., Delaney, T.P., Lazarowitz, S.G., Martin, G.B., Schneider, D.J., Tang, X., Bender, C.L., White, O., Fraser, C.M. and Collier, A.			
The complete genome sequence of the Arabidopsis and tomato pathogen Pseudomonas syringae pv. tomato DC3000			
Proc. Natl. Acad. Sci. U.S.A. 100 (18), 10181-10186 (2003)			
2 (bases 1 to 155628)			
Buell, C.R., Joardar, V., Khouri, H., Fedorova, N., Tran, B., Russell, D., Berry, K., Uterback, T., Van Aken, S., Feldblyum, T., Gwyn, M., Dodson, R., Deboy, R., Durkin, A., Kolonay, J., Madupu, R., Daugherty, S., Brinkac, L., Beanan, M., Haft, D., Selengut, J., Nelson, W., Davidson, T., White, O., Fraser, C. and Collier, A.			
Direct Submission			
Submitted (03-MAR-2003) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA			
Location/Qualifiers			
1..155628			
/organism="Pseudomonas syringae pv. tomato str. DC3000"			
/mol_type="genomic DNA"			
/strain="DC3000"			
/db_xref="taxon:223283"			
/note="pathovar: tomato"			
complement(159..4811)			
/locus_tag="PSPT05481"			
complement(159..4811)			
/locus_tag="PSPT05481"			
/note="identified by Glimmer2; putative"			
/codon_start=1			
/transl_table=1			
/product="hypothetical protein"			
/protein_id="AAO58900.1"			
/db_xref="GI:28855842"			
/translation="MTSTPLQSNVHTAVAPAPASQQRSDLGIFCOALVGLTPVGEYH DKLANSILNOOLAPEROIAIHQLAIPVSHWDSNIHPTDPTAMALAGAEYLRM AIIHVPDGAWNTAIEDATVAPAGMALSAISIGEHRSLESILAAIYLMVY LHSAVALDESLPNKSLIDAPRDPDSTVYRGLIQRGMGLCPESGSLFRS PHATVHALCCTPAQFKSVFKPLSLTMDWFGSGSGHTSPVLTQALAGVYADHVAAYQ			



```

LDGSLRKALRRKRVCSHSHQVHDKVKRKYRAFPQALPSTLDIMCHLPIREAMPBL
LVEGVPHLQYKRSQSVAFTHGVALVBLSPGLSQITPTDTLLKVSADLQSSDVID
HAAWAKTLIIPALRYATAGAIEMWKNNDVQAASQISQALATLDQOSHAQELNS
LIGIKPADRKGLAQOOLKAGVDHMLWQSIHIDHWPLIQAHGFTVASYSIDTLAA
GRPGASVELVMEGEVYIQGQVPEAVATAPDAFORALVBAEASVIRLOLSEQVID
OKTLLDSTCELSRFGAOSTGGLFRCOAGDRSRBEGHVEGERPELPIAGVAG
EVRAAPRYABEGVTMSGTIPITTEALSKEAKORIREKARPTPLMPNDSADLCAVSR
SSAALHPQOQSTLPSATLVFADTDOQSGLDVLAASKAEHLGDFLERSVAEHLHT
QMEVMAKEREMADIARLLIPFYGCIKDLASGRSGGAIIVGCVMDVAFLVLPQGVY
SSTARIVLRAGMESVESARLIGQAVGLIAGLAKSSAVFALRDIIGRLGKAGDGA
ALBEDVSLSSVSSQAVLDATFGLDGMVRLADDMOLPTAAPSAPVMDGADVY
VRDGTLEQDPDFRLSDPSGGAHGGKLTQVNASRMEVSVAOERSPEHYPSIAAY
TVEVGTACEVRIYERCNVRIEREGVFIDLVDQIYHIDANTPDALARKLAVEKLS
LRSTQLETERLCKVRNRLIRI PCTNGVKTALPAEPIMDGSTPKRTGXPSNAMA
REBELDRLSLDADSAASIDVFVNEGKCKMADSEVEPAASAQVQVPLSEARML
LALPETYVLEPOLGALRAGLGLPENFPMDDALMIYAHIPVIGIPLASIDADA
LGRIRLEMDGNMTFIEPDGTFYKAPISMDAEDLKRSRYTEAAEINEYRVEQYR
LVREPGABODQENIARLLPDLLDSARADHNSWGRVTHYDYVOMCRANQPNL
LKPRANMSGSEIQKFTVTLARNSTIPDKITLRSLPQOCHI VAILNQLPTQSPYIK
WEKTLTSTIVTPQAPKRIKQVKGANISFLQATYESGERIVYVALSGNNAKDKQLD
DVTESTERVIDGVYRDARARAMAGROPFGTSLPVIRVDVHLVRSFGRXLDSERLI
ATVLEKDMASTFKLHIKVFVTLDTCRSCGCVLPRLKLDLPDAQFSVTYLYKPYQAI"
complement (5035..5928)
/locus_tag="PSPOT5482"
complement (5035..5928)
/locus_tag="PSPOT5482"
/codon_start=1
/transl_table=11
/product="response regulator"
/protein_id="AA058901.1"
/db_xref="GI:28855843"
/translacion="MSKSVYLVDVDAFTRIDLVKQGLRNYEPQIHTEDAVNGRAQAL
LSKESPLILCDMEPMESGLIELTWCEODNYLKTVPFIWTSRGKENVVQALQAG
LTDVVGQPTNTEBOLSKYKALAVGKLDAMSTAPRANSPINDLSALTGVAEYLR
RATSAAPASITPKAPPYAAVARSAPETRGQGLALPNGTQPCVKALSKDATTIVAY
SEKLPQVLESADVLEDQESNEVARLNGYLSVAFAEQKPUSEMLQVTFRRVDDAK
LDYLSRLIARQTAQKHFVPGA"
complement (6131..6892)
/gene="phou"
/locus_tag="PSPOT5483"
complement (6131..6892)
/gene="phou"
/locus_tag="PSPOT5483"
/codon_start=1
/transl_table=11
/product="phosphate transport system protein Phou"
/protein_id="AA058902.1"
/db_xref="GI:28855844"
/translacion="MIKHSHTHHISQOFNALEVDVSHLLEMGLEKQVNDSTAL
IEADSGIARVREVDRIINOMERSIDECLRIARROPASDIALIISISKVIDLER
IGBATYIARBAIKLDCGEAPRGVYVRHIGQVRRMVWDALDSFAPARELALSAV
OYDKITIREYKTALRELATYMEDPRISRYLDVIYVLRSLERIGHARNISELVITL
VRGTDVHMGKKEKEVQGISSTANVPGNADK"
complement (6963..7796)
/gene="pscB-2"
/locus_tag="PSPOT5484"
complement (6963..7796)
/gene="pscB-2"
/locus_tag="PSPOT5484"
/codon_start=1
/transl_table=11
/product="phosphate ABC transporter, ATP-binding protein"
/protein_id="AA058903.1"
/db_xref="GI:28855845"
/translacion="MOHERPTHGNINIPALRRKQVLDLAETVAIEVPGSLFYGDKQ
ALPDIANIPKQVTSFISGSCGCKSTLLSFNNMNDLVGCRVEGAINVGNHNYTK
GEVAVLRRRGVFOKPNPEPKTIYENVVGLRIGINKKRVDEAVANAKKAAALYK
DEVDRLEHESALGSGGQOORIVARTIAREPEVLLIDPCSNALDIPSTIKVEELIYE
LKSXTYIVTYHNNQAAARVSDYTAFFMTMGLVRFPGTDITLFTNPAKQIBDYITGXY
G"
complement (7863..9536)
/locus_tag="PSPOT5485"
complement (7863..9536)

```

```

/locus_tag="PSPOT5485"
/codon_start=1
/transl_table=11
/product="phosphate ABC transporter, permease protein,
putative"
/protein_id="AA058904.1"
/db_xref="GI:28855846"
/translacion="MLKRNLSKMFPSKSGAGVMSAGAVSAIVIMTIGLLAVIARGL
GHMVPADLVAAQVAPDQPAHVIVGELVBOQVPRARLAKAGLPVPOQGEFTRELF
KNGNRDLPNDPFTVWVEMLSGSSPAELMTLRRERENGFNFYLVNVEGGRVVAEGE
NAMPYLOERIKRLETTADELXSLKQDIGAINGIDBELRLOARKLELAGLDAADQD
MAARBAELDRKRYTIBERLETHQAPRDSLTARDADGKTEISLGVVAAQPNANG
TFKLGFFYQKLEFSDDPREKNTBEGCPALFPAIVGVTLLNAVIVTPRGVLAATL
REARQOQVYRLRIANVNLAVGPAVIVYGVGFFVYVIGSVDRLPFPBALPAPTC
GETGLADSLTLLALVAVVIVATEEBELARIPTLREBSALGATKAETLMKTLVPA
SPAMNTQILIAVRAAGEVAPLMVGVVKAAPLPVDQGVNPLYHLNOKIMHGFHVID
VGSQPVVEARPLVAVATALLVIALNLAVSIRNHLREKYLKLDN"
complement (9548..11581)
/locus_tag="PSPOT5486"
complement (9548..11581)
/locus_tag="PSPOT5486"
/codon_start=1
/transl_table=11
/product="phosphate ABC transporter, permease protein,
putative"
/protein_id="AA058905.1"
/db_xref="GI:28855847"
/translacion="MOGAGEPLMISTIEONLIGMRVSDKGEVIFPNVQGAELHRVAL
PIRAGSVASIGDQGRAPILNAGLSNGOYLIRHRYLTTPYNNQKITTSVAMPYBE
TPVLVDAGRAVSHVTVSADGSLKVAQSGTGLDHWLSLQKNNMTGENTRODTRIE
LPQMSAEVKAIVYIDPQOQMYVINGRAQADVPSLRSLNGARKLLEDGAETASQO
LVNGISLILNNSGSSQOMFARMDTQGEQFMSIRPQKQKAPRQIRARQRRGFPA
LDSGRGKGVHSAHRLIYKVAADGGILTLPBANQVLESGGKLPULLNDPHE
VSSSLMGKGVSHARLIVKVAADGGILTLPBANQVLESGGKLPULLNDPHE
AVAAIYTTAYFMAPKRRKRYVIELMEAMPYIILGFPAFLPAPYIEALPGIPLSL
LLPPIGILLAGFWTRLPDSIRLRVDPQWEGALILPVLVWGPSLSMSPLHNWFG
GDMSTYLRDELGITDORNALVVGANGVAFINIVISLADAVFSVRSLLTGLSALG
ATPMQTLTRVILITASPGIFSAIMIGKRAVGETMIVLWNTGPTIINDNMLFGMRTL
NANVAVEMPESEVGGSHYRVLFLAALVLLFTFVMTLAILNQRILKQVSL"
complement (11962..12960)
/locus_tag="PSPOT5487"
complement (11962..12960)
/locus_tag="PSPOT5487"
/codon_start=1
/transl_table=11
/product="phosphate ABC transporter, periplasmic
phosphate-binding protein, putative"
/protein_id="AA058906.1"
/db_xref="GI:28855848"
/translacion="MTLKRIMTALTPVAGVATHTTPAADVPAIKAVYTKTSGVNSLS
SVSDPLANMTIMAAKRYKESVNIQIOAASSSTRAPALVTGTANTGMSRBMQD
ELSAFEQKHGKRTPAIIVAVDALAVFHKQNPILKGLILOVDVAFSTRICGAKADK
TWDVGVYDGLACKPIQLFGRNSVSGTYGFKFEKALCKQPKANVQEPQSSAVSVAI
SNSINGISGIGCYKTSNVRTPLAKKAGGEPEDBANLNTGYPLSRPLYVYNVA
PNOPLALEAFKVLVLSROGQEVNMKGVIPLPARVEKTLTDLGLQSGASGLAVK"
complement (13130..14467)
/locus_tag="PSPOT5488"
complement (13130..14467)
/locus_tag="PSPOT5488"
/codon_start=1
/transl_table=11
/product="major facilitator family transporter"
/protein_id="AA058907.1"
/db_xref="GI:28855849"
/translacion="MUSLAPFEQVRSLSVMSVSSSTOPARPLTRSDYKTLSSNL
GGLAEVYDFIIFVFAVVGKLEFPVDMQDMLEMOQTFGIFAAGYIARPGCITIMHF
GAVFGEHQPWETLSIFMAAVPTLIMGLPTYAOIGMAVAPILLVWVIOGAAIGSEVQ
AMFVSEHVAARVAVACGTTAGTILGILSLVATATNSITPVEVSEYARIRPIL
LGVGFMSYVLRRLIHERPVPABEIDURKALAEVYKAVLRHRGAVALSHMTMLL
SAGIIVILMTFTIITQITIGFPAATIALKANSLAIVLSGCCIAAGLADRFGAGVFL
LSAGAGLVSMTFVHVGSHPEVLPVYVITGIFVGTIGAVPVMVAKAPPPVYRFGSL
SFSYNLAIVAFGLTPMVVTFMLKSSPMGSPSVYVALICGIMAGIVGLLTRK"

```

Query Match	95.6%;	Score 17.2;	DB 1;	Length 155628;
Best Local Similarity	88.9%;	Fred. No. 5.6e+02;		
Matches 16;	Conservative 2;	Mismatches 0;	Indels 0;	Gaps 0;
Db	24486	GACCACGTGCGGCATG 24503		
Oy	1	GACCACGTGCGGCATG 18		
RESULT 4				
AP005035/c		296300 bp	DNA	linear
LOCUS				BCT 10-MAY-2003
DEFINITION	AP005035			Streptomyces avermectilis genomic DNA, complete genome, section
ACCESSION	AP005035	BA000030		
VERSION	AP005035.1	GI:29607035		
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
MEDLINE				
PUBMED				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				

CDS  
DESGRLGSESG"  
5283..7997  
/note="SAV3388"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAC71100.1"  
/db\_xref="GI:29607041"  
/translation="MGYLSRRKTHAYPLSLKRRTWLTGAVAGAVVTAVASPS  
GDADVDVARGRPVEVHLAKAGTSRRLARTTAFSLVGVSMGARELGTQ  
VRRTGTGEGSWRNLAFLPVDRPEGANAGASDPLVGPSPDQARVLADS  
ARAGLPGLEVNIVDPVTTTARRNKGLDSTGTGDI.SNAAVADPSOTGTS  
GSATGSEPARAPGSDSASATATATATATTA.PTAPASVAPRSVSTRMADS  
AVAGSPQYIDRISAVFVHTRASNDYSCQASALVIRIMADIVARGLDGINFLV  
KCGRI.FEGRAGGADL.PVRGHTYGFNDSTGI.AVLGPEGSAAAPKPSRAESVA  
RLAAMKLGVOYGNPSTVTLTADADTVGVAADATNLNLSGGKDAATTS.PGKSLY  
KLGEVRVYASPGSSAI.PTADYNGDVSDLVATPKQSGMVLTVPGSLGPGVSAK  
LKNGSTGVPGAESGDQMGATAMGDINGDVAADLAVAGREDDTHADRGVTL  
YGPKEFDGADTMALGDDYNNRSPAFGATVAVGDNRADGKADVTATATGTGNVAFRA  
NGQETADITTVSGALVADASGDFNRDGVADVALTYRDSGVGRVTFKSGALGL  
SKVSTLVKGRSLAADVNGDYDVI.GQPSASESGSGGQVTLVPGASTGFTT  
GMTVHGTTGVEGASESDAFCTSVSGVDNADGVADVLGAPSEDIITDARKRNA  
GSVWLKGAASSGLTGTSLASDTANI.PGSTETDDKGLSLSDVLTGGGVADLTIG  
ABGDDTCTGLVYPTVGTGVTAKAYYGAQLGISTGRLGQLTP"  
complement(8074..8532)  
/note="SAV3389"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAC71101.1"  
/db\_xref="GI:29607042"  
/translation="MGRSCAGVGNPSTGRDESWERAPGVNVTETVRPOAAKPVW  
AADAVAMRSGARLIDYSQSLWRVMEIGIRGAPYAAVAEAVARGGAYGEV  
VRHTGEMTTGGDHVWRTPDGRMDPIDEARCYLDGSLRLCRDAAD"  
complement(8607..9167)  
/note="SAV3390"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAC71102.1"  
/db\_xref="GI:29607043"  
/translation="MSLVELEADDERGLAASGLACIDRCVPLLGDDVLRPLKASL  
VEGEWERLTKAEALGADAEPTAEDECGIARMLVAPQPAAGALREKADVC  
SVAAQIHLRLDADDPGVSARBERGTGMSPLVAELRRQVTVLELVAGAGCGL  
RRALDVTEGRVLRVAVSRARRGRA"  
complement(9252..9779)  
/note="SAV3391"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAC71103.1"  
/db\_xref="GI:29607044"  
/translation="WATWGYFVGRSRPLAELGALTAAAGMTLLQTAPOGOWEYF  
SGDDVSGMNALETAGAPALFGVNNVSCVAAAPQAGMTTCIARMAAYLNG  
GPGDEQEGVEDEYPLTPGEAERAAVAAAEAGHTVAPGLAQVLGVDDPPEPLAE  
NLFRFLRLGVEPL"  
complement(9880..10941)  
/note="SAV3392"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAC71104.1"  
/db\_xref="GI:29607045"  
/translation="MDMRFVPCVRFGEERVAFTDTPGRGRTMLCTRCRHAASRP  
BGLCTCANGTTSTVPAFDAAAGAPQAPAPSPDPDAGFQSDGPGAPAPGGG

Query Match 95.6%; Score 17.2; DB 1; Length 296300;  
Best Local Similarity 88.9%; Pred. No. 5e+02;  
Matches 16; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

CDS  
1 GACCACGTSCCGGCATG 18  
Db 157263 GACCACGTSCCGGCATG 157246

RESULT 5  
AB001608  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
gene  
CDS

AB001608  
Streptomyces virginiae DNA for BarX, complete cde.  
AB001608  
AB001608.1 GI:2641955  
BarX.  
Streptomyces virginiae  
Streptomyces virginiae  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
1 (sites)  
Kinoshita, H., Ipposhi, H., Okamoto, S., Nakano, H., Nihira, T. and  
Yamada, Y.  
Bacterial autoregulator receptor protein (BarX) as a  
transcriptional regulator in Streptomyces virginiae  
J. Bacteriol. 179 (22), 6986-6993 (1997)  
98037495  
9371444  
2 (bases 1 to 1215)  
Kinoshita, H.  
Direct Submission  
Submitted (06-MAR-1997) Hiroshi Kinoshita, Osaka University,  
Department of Engineering; Yamadaoka 2-1, Suita, Osaka 565, Japan  
(E-mail:kinoshita@biochem.bio.eng.osaka-u.ac.jp, Tel:+81-6-879-7433,  
Fax:+81-6-879-7432)  
Location/Qualifiers  
1..1215  
/organism="Streptomyces virginiae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:1961"  
244..1128  
/gene="barX"  
244..1128  
/gene="barX"  
/codon\_start=1  
/transl\_table=1  
/product="BarX"  
/protein\_id="BA023611.1"  
/db\_xref="GI:2641956"  
/translation="MTSTVRELVHRAVAEVLFTGMSRTAENRPLATTAQWPAHSYF  
TPNGCYDPLALSETIRVOGTLSHAFFGVSPEDQPLMDLHNSVPEQGVGAAPAD  
LELDVLCSDIRRRGRRLAGKRYEVLVCGGVATGGAAPDCTSPAYORLDRVGA  
TGVRLP.PQPLAPASVGRFLTTDVLSATEPLEMOULRVDOHVLFDHPVDVHGMVL  
MESARQAAQIDSRPPLPTTMASERSRAELDRPCWIAEPLPADNGRQVATVGH  
QDDTVFSCLTGTRGAAE"

ORIGIN  
Query Match 88.9%; Score 16; DB 1; Length 1215;  
Best Local Similarity 88.9%; Pred. No. 5.1e+03;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Cy 1 GACCACGTSCCGGCATG 18  
Db 874 GACCACGTSCCGGCATG 891

RESULT 6  
AB001683  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

AB001683  
Streptomyces sp. gene for FarX, FarX, complete cds.  
AB001683  
AB001683.1 GI:2342428  
FarX.  
Streptomyces sp.  
Streptomyces sp.  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
1 (sites)  
Waki, M., Nihira, T. and Yamada, Y.  
Cloning and characterization of the gene (farX) encoding the

receptor for an extracellular regulatory factor (IM-2) from  
Streptomyces sp. strain FRI-5  
J. Bacteriol. 179 (16), 5131-5137 (1997)

JOURNAL  
MEDLINE  
97405912  
PUBMED  
2 (bases 1 to 2234)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Waki,M.  
Direct Submission  
Submitted (07-MAR-1997) Miyoko Waki, Osaka University, Graduate  
School of Engineering, Department of Biotechnology; Yamadaoka 2-1,  
Suta, Osaka 565, Japan (E-mail:waki.yam@stu.bio.eng.osaka-u.ac.jp.  
Tel:+81-6-879-7433, Fax:+81-6-879-7432)

FEATURES  
source  
location/Qualifiers  
1..2234  
/organism="Streptomyces sp."  
/mol\_type="genomic DNA"  
/strain="FRI-5"  
/db\_xref="taxon:1931"  
440..1315  
/gene="farX"  
440..1315  
/gene="farX"  
/codon\_start=1  
/transl\_except=(pos:440..442,aa:Met)  
/transl\_table=11  
/product="PairX"  
/protein\_id="BAA21858.1"  
/db\_xref="GI:2342429"  
/translation="MWHTSTAOVLTIDMRLDPAFSVTARWPLSHAFPPVVDGY  
DPLCAETRIOTAILGHAERFVPCGHFVMDISVSRPELLRVLGVPATVLAIT  
CVELRRARISGSLYEAVVRDQVAVTGAATCTSPAYTORIPEHVTTPPEHPL  
PLTAPAPQSVARLSPDVLSPLDRERWQLRVDTNHPVLFDMVDHVPKVMLEAA  
ROAASALGRPSFMPLGVAGGFKRYVELDAPCVIESERLFDVVGAEVAVTGHONG  
ELTFVGVTASSYQ"  
1500..2165  
/gene="farA"  
1500..2165  
/gene="farA"  
/codon\_start=1  
/transl\_except=(pos:1500..1502,aa:Met)  
/transl\_table=11  
/product="PairA"  
/protein\_id="BAA21859.1"  
/db\_xref="GI:2342430"  
/translation="MAEQVRAIRTEQALISAARVFDERGVQAATISILLTAVGTG  
ALYFPOKEDIAQGLTANONEDLLPERPAQLOEVDAVWMLHTHRLTNPVAGVR  
LSIDVNAAGLDRSAPRWWDKFTDLEKQAQGLPHVYPAETADVTGAVGVOS  
MSQALTERHODIGQRVNALRLHMLPSIAQPSVLASLHGESRAEEVYLEARQIARQAD  
EED"

ORIGIN  
Query Match 88.9%; Score 16; DB 1; Length 2234;  
Beet Local Similarity 88.9%; Pred.No. 4,6e+03;  
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

CY 1 GACCACGTCSCGGCGATG 18  
|||||  
DB 1056 GACCACTGACCGGCATG 1075

RESULT 7  
SC0007731 6727 bp DNA linear BCT 24-AUG-1998  
LOCUS 07731/c  
DEFINITION Streptomyces coelicolor scbR gene, scbA gene, ORFs A,B,X & Z.  
ACCESSION AJ007731.1 GI:3425857  
VERSION  
KEYWORDS gamma-butyrolactone binding protein; scbA gene; scbR gene.  
SOURCE Streptomyces coelicolor A3(2)  
ORGANISM Streptomyces coelicolor A3(2)  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Streptomycineae; Streptomycetaceae; Streptomyces.  
REFERENCE  
1 Takano,E., Chakraborty,R., Nihira,T., Yamada,Y. and Bibb,M.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (11-ANG-1998) Takano E., Dept. Genetics, John Innes  
Institute, Colney lane, Norwich, NR4 7UH, UK

FEATURES  
source  
location/Qualifiers  
1..6727  
/organism="Streptomyces coelicolor A3(2)"  
/mol\_type="genomic DNA"  
/strain="M145"  
/db\_xref="taxon:100226"  
37..969  
/gene="orfB"  
37..969  
/gene="orfB"  
/codon\_start=1  
/transl\_table=1  
/product="histidine kinase"  
/protein\_id="CAA07625.1"  
/db\_xref="GI:3425858"  
/db\_xref="GOA:O86849"  
/db\_xref="UniProt/TREMBL:O86849"  
/translation="MGTPSPTEENRGGLIPCSIDRPVSSRRTTERHPCRPTRPHD  
LHAGRHGRGRTETLQDGLSRSPRRARQITGTGTHPGIPAPGRPPAGPVTDRGRD  
HVMHERRRLAREVDELTGALSTPAHHI ELHAAETGCTPLDAAHSLRAEMVARRL  
TGELQAOQTVPLPLAQAAGEPAAITRPGTGVRLRTTGDESLDVCRESLEFLVREAL  
HNAFHNAHARVYTLTAREPRMAAGIVDQGVCPDAVALVPGHRAGLRSMRDIED  
VGMRLTVSGAGGTHIDVLPRLPRKYSTAPRT"  
complement (1000..1854)  
/gene="orfA"  
complement (1000..1854)  
/gene="orfA"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="CAA07626.1"  
/db\_xref="GI:3425859"  
/db\_xref="UniProt/TREMBL:O86850"  
/translation="MLIVTGYTVGQVVAHLARTGPVRLARRPRVYTGPNVEY  
CAGSYADRFGLDRALNGVRAAFVLTNSATPDDRFAPAAAREGVRHLVKLSMLAVE  
PGADDFITRQREVERAIVRSGLDWTFLRARTFMSNTLSMAPAIRSGVVRALYGTSP  
VACVDPDIEVAVRALTRPGHEBRAVALGPEALITRAOQTARISEVLGSGSLRPEEG  
LEARBEOMLRKYPVADAFLESAERQAGAKQAVPTVAVETGRRPARPRIWADHA  
EARGSQG"  
complement (1970..2914)  
/gene="scbA"  
complement (1970..2914)  
/gene="scbA"  
/function="involved in gamma-butyrolactone synthesis"  
/codon\_start=1  
/transl\_table=1  
/product="ScbA protein"  
/protein\_id="CAA07627.1"  
/db\_xref="GI:3425860"  
/db\_xref="UniProt/TREMBL:O86851"  
/translation="MPEAVVLINSASANSISLQALRPVPMALVHRTVQDAFPYSWTP  
KGGDRFSVTAVLPDHPFPFVHGDRDPLLIATLLOAMLVFHAIGFVGHFLM  
ATLDYTLCHLDLGVSGEVALEVEVAQSQKFRGQPVQGVDAVVRAGLAAATGTA  
TTTRFSPQVYRMRGDAITPTASVGPAPVPAAGRTDIEDVLTSSASQODTWRLEV  
DTSRPTLFORNDHVPGLMLLEAARQACLVTPGAPVPSIGTRFVRYAEFSDPCWI  
QATVTPGAPAGLTIVRYTGHQDGLVFLTLISGAPFG"  
3032..3679  
/gene="scbR"  
3032..3679  
/gene="scbR"  
/codon\_start=1  
/transl\_table=1  
/product="gamma-butyrolactone binding protein"



```

gene
3042. .4190
/gene="SC06260"
/note="synonym: SCAH10.25"
3042. .4190
/gene="SC06260"
/note="SCAH10.25, possible sugar kinase, len: 382 aa; similar to SW:GLK_STRCO (EMBL:X65932) Streptomyces coelicolor glucokinase (EC 2.7.1.2) (glucose kinase) Glk, 317 aa; fasta scores: opt: 1022 z-score: 1142.9 E(): 0; 46.6% identity in 311 aa overlap and to TR:CAB51974 EMBL:SC6E10 Streptomyces coelicolor SC6E10.20c, 317 aa; fasta scores: opt: 1022 z-score: 1035.4 E(): 0; 46.6% identity in 311 aa overlap. Contains a match to Pfam entry PF00480 ROK, ROK family and Prosite entry PS01125 ROK family signature"
/codon_start=1
/translation=11
/product="putative sugar kinase"
/protein_id="CAB60179.1"
/db_xref="GI:6273665"
/db_xref="GOA:Q9RKT2"
/db_xref="UniProt/TrEMBL:Q9RKT2"
/translation="MSYTRDFTAPIGSGRRAPVLRVTGTRRSHLTAPRPVTVGIDIG
GTYKAGVVDADGNITLKTETPDKSKSVVEDITVELVLDSDRHDAVAGIAA
GWDADRNRYLPAPLSWRNEPLRDRAGLAAPVLDVNDPANTAAEMFEGSGSD
HIVITLGTGIGAILLEDQGVKRGKVAPEGHMGPVPGHRCPCNRCGREGSDG
NALVREBELAANDSPVAIVGIIIEHVKSIDITDTPMTTELAREBDANCVELDIOGM
LVGINALAAALDPSLCVIGGSVADDLIGPARDFKQLTGRIPEARIVRAQL
GPGAGWGADLPLVARFRRAKRRRVERERYERYAEARRESRL"
3159. .3704
/gene="SC06260"
/note="Pfam match to entry PF00480 ROK, ROK family, score 215.50, E-value 5.4e-63"
3540. .3623
/gene="SC06260"
/note="PS01125 ROK family signature"
4244. .4247
4259. .4768
/gene="SC06261"

Query Match 88.9%; Score 16; DB 1; Length 290850;
Best Local Similarity 88.9%; Pred. No. 1.8e+03;
Matches 16; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCACGTCSCCGGCATG 18
Dbb 12783 GACCACGTCACCGGCATG 12800

RESULT 9
BD139044 513 bp DNA linear PAT 18-SEP-2002
LOCUS BD139044
DEFINITION Virulence-associated nucleic acid sequences and uses thereof.
ACCESSION BD139044
VERSION BD139044.1 GI:23233989
KEYWORDS JP 20020505849-A/92.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria: Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 513)
Aasubel, F., Goodman, H.M., Rahme, L.G., Miklos, S.M., Tan, M.W.,
Cao, H., Drenthard, E. and Tsongalis, J.
Virulence-associated nucleic acid sequences and uses thereof
Patent: JP 20020505849-A 92 26-FEB-2002;
THE GENERAL HOSPITAL CORP
OS Pseudomonas aeruginosa
PN JP 20020505849-A/92
PD 26-FEB-2002
PF 25-NOV-1998 JP 2000522270
PR 25-NOV-1997 US 60/065517
PI FREDERICK AUSUBEL, HOWARD M GOODMAN, LAURENCE G RAHME PI
, SHALINA MAHAJAN MIKLOS,

```

PI MAN WAH TAN, HUI CAO, ELIANA DRENKARD, JOHN TSONGALIS, PC  
 C12N15/09, A61K45/00, A61P31/04, C07K14/21, C12Q1/02, C12Q1/68, PC  
 G01N33/15,  
 PC G01N33/50//C12P21/02, (C12N15/09, C12R1:385), C12N15/00, (C12N15/  
 PC 00, C12R1:385)  
 CC Virulence-associated nucleic acid sequences and uses thereof  
 FH Key Location/Qualifiers  
 FT source 1..513  
 /organism='Pseudomonas aeruginosa'.  
 /mol\_type='genomic DNA'  
 /db\_xref='taxon:287'

FEATURES  
 source  
 1..513  
 /organism='Pseudomonas aeruginosa'

ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 513;  
 Best Local Similarity 83.3%; Pred. No. 9.3e+03;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACTGTCGGCGCATG 18  
 |||||:|:|:|:|:|:|  
 133 GACCACTGTCGGCGCATG 150

RESULT 10  
 ARI98867 513 bp DNA linear PAT 20-APR-2002  
 LOCUS  
 DEFINITION Sequence 155 from patent US 6355411.  
 ARI98867  
 VERSION ARI98867.1 GI:20248941  
 KEYWORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE  
 1 (bases 1 to 513)  
 Ausubel, F., Goodman, H.M., Rahne, L.G., Mahajan-Miklos, S., Tan, M.-W.,  
 Cao, H., Drenkard, E. and Tsongalis, J.  
 Virulence-associated nucleic acid sequences and uses thereof  
 JOURNAL Patent: US 6355411-A 155 12-MAR-2002;  
 FEATURES  
 source  
 1..513  
 /organism='unknown'  
 /mol\_type='unassigned DNA'

ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 513;  
 Best Local Similarity 83.3%; Pred. No. 9.3e+03;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACTGTCGGCGCATG 18  
 |||||:|:|:|:|:|:|  
 133 GACCACTGTCGGCGCATG 150

RESULT 11  
 CQ585804 744 bp DNA linear PAT 02-FEB-2004  
 LOCUS  
 DEFINITION Sequence 13562 from Patent WO0171042.  
 CQ585804  
 ACCESSION CQ585804.1 GI:41645849  
 VERSION  
 KEYWORDS  
 SOURCE Drosophila sp.  
 ORGANISM Drosophila sp.  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 REFERENCE  
 1 Venter, J.C., Adams, M., Li, P.W. and Myers, E.W.  
 Detection kits, such as nucleic acid arrays, for detecting the  
 expression of 10,000 or more Drosophila genes and uses thereof  
 Patent: WO 0171042-A 13562 27-SEP-2001;  
 JOURNAL PE Corporation (NY) (US)

FEATURES  
 source  
 1..744  
 /organism='Drosophila sp.'  
 /mol\_type='unassigned DNA'  
 /db\_xref='taxon:7242'

ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 744;  
 Best Local Similarity 83.3%; Pred. No. 8.6e+03;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACTGTCGGCGCATG 18  
 |||||:|:|:|:|:|:|  
 Db 454 GTCACAGTCCGCGCATG 471

RESULT 12  
 BD139042/c 762 bp DNA linear PAT 18-SEP-2002  
 LOCUS  
 DEFINITION Virulence-associated nucleic acid sequences and uses thereof.  
 BD139042  
 ACCESSION BD139042.1 GI:23233987  
 VERSION BD139042.1 GI:23233987  
 KEYWORDS JP 2002505849-A/90.  
 SOURCE Pseudomonas aeruginosa  
 ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.  
 REFERENCE  
 1 (bases 1 to 762)  
 Ausubel, F., Goodman, H.M., Rahne, L.G., Miklos, S.M., Tan, M.W.,  
 Cao, H., Drenkard, E. and Tsongalis, J.  
 Virulence-associated nucleic acid sequences and uses thereof  
 JOURNAL Patent: JP 2002505849-A 90 26-FEB-2002;  
 THE GENERAL HOSPITAL CORP  
 OS Pseudomonas aeruginosa  
 PN JP 2002505849-A/90  
 PD 26-FEB-2002 JP 2000522270  
 PF 25-NOV-1998 US 60/066517  
 PR 25-NOV-1997 US 60/066517  
 PI FREDERICK AUSUBEL, HOWARD M GOODMAN, LAURENCE G RAHNE PI  
 , SHALINA MAHAJAN MIKLOS,  
 PI MAN WAH TAN, HUI CAO, ELIANA DRENKARD, JOHN TSONGALIS PC  
 C12N15/09, A61K45/00, A61P31/04, C07K14/21, C12Q1/02, C12Q1/68, PC  
 G01N33/15,  
 PC G01N33/50//C12P21/02, (C12N15/09, C12R1:385), C12N15/00, (C12N15/  
 PC 00, C12R1:385)  
 CC Virulence-associated nucleic acid sequences and uses thereof  
 CC Virulence-associated nucleic acid sequences and uses thereof  
 FH Key Location/Qualifiers  
 FT source 1..762  
 /organism='Pseudomonas aeruginosa'.  
 /mol\_type='genomic DNA'  
 /db\_xref='taxon:287'

FEATURES  
 source  
 1..762  
 /organism='Pseudomonas aeruginosa'  
 /mol\_type='genomic DNA'  
 /db\_xref='taxon:287'

ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 762;  
 Best Local Similarity 83.3%; Pred. No. 8.6e+03;  
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACTGTCGGCGCATG 18  
 |||||:|:|:|:|:|:|  
 Db 303 GACCACTGTCGGCGCATG 286

RESULT 13  
 ARI98865/c 762 bp DNA linear PAT 20-APR-2002  
 LOCUS  
 DEFINITION Sequence 153 from patent US 6355411.  
 ARI98865  
 ACCESSION ARI98865.1 GI:20248939  
 VERSION  
 KEYWORDS  
 SOURCE Unknown.



```

ORGANISM      Unknown.
REFERENCE     Unclassified.
AUTHORS       1 (bases 1 to 762)
              Aueubel,F., Goodman,H.M., Rahne,L.G., Mahajan-Miklos,S., Tan,M.-W.,
              Cao,H., Drenthard,A. and Tsongalis,J.
              Vitulence-associated nucleic acid sequences and uses thereof
              Patent: US 6355411-A 153 12-MAR-2002;
FEATURES      location/Qualifiers
              1..762
              /organism="unknown"
              /mol_type="unassigned DNA"

ORIGIN
Query Match      86.7%  Score 15.6;  DB 6;  Length 762;
Best Local Similarity 83.3%  Pred. No. 8.6e+03;
Matches 15;  Conservative 2;  Mismatches 1;  Indels 0;  Gaps 0;

QY
1 GACCACGTCTCCGGGCATG 18
|||||:|:|:|
Db 303 GACCACGTCTCCGGGCATG 286

RESULT 14
AY060758      859 bp  mRNA  linear  INV 08-NOV-2001
LOCUS         Drosophila melanogaster GH20994 full length cDNA.
DEFINITION   AY060758
ACCESSION    AY060758.1 GI:16768173
VERSION      F11_CDNA.
KEYWORDS     Drosophila melanogaster (fruit fly)
SOURCE       Drosophila melanogaster
ORGANISM     Drosophila melanogaster
              Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
              Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
              Ephydroidea; Drosophilidae; Drosophila.
              1 (bases 1 to 859)
REFERENCE    Stapleton,M., Bickstein,P., Hong,L., Asghyani,A., Carlson,J.,
AUTHORS      Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,B., George,R.,
              Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,
              Nunoo,J., Pateb,J., Paragas,V., Park,S., Phouenavong,S., Wan,K.,
              Yu,C., Lewis,S.E., Rubin,G.M. and Ceiniker,S.
              Direct Submission
              Submitted (30-OCT-2001) Berkeley Drosophila Genome Project,
              Lawrence Berkeley National Laboratory, One Cyclotron Road,
              Berkeley, CA 94720, USA
COMMENT      Sequence submitted by:
              Berkeley Drosophila Genome Project
              Lawrence Berkeley National Laboratory
              Berkeley, CA 94720
              This clone was sequenced as part of a high-throughput process to
              sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
              Science 2000). The sequence has been subjected to integrity checks
              for sequence accuracy, presence of a polyA tail and contiguity
              within 100 kb in the genome. Thus we believe the sequence to
              reflect accurately this particular cDNA clone. However, there are
              artifacts associated with the generation of cDNA clones that may
              have not been detected in our initial analyses such as internal
              priming, priming from contaminating genomic DNA, retained introns
              due to reverse transcription of unspliced precursor RNAs, and
              reverse transcriptase errors that result in single base changes.
              For further information about this sequence, including its location
              and relationship to other sequences, please visit our Web site
              (http://fruitfly.berkeley.edu) or send email to
              cdna@fruitfly.berkeley.edu.
              location/Qualifiers
              1..859
              /organism="Drosophila melanogaster"
              /mol_type="mRNA"
              /strain="y; cn bw sp"
              /db_xref="taxon:7227"
              /map="50C9-50C11"
              1..859
              /gene="CG6305"
              /note="alignment with genomic scaffold AB003817"

```

	CDS	/db xref="FLYBASE:Fbgm0033869"
		208..744
		/gene="CG6305"
		/note="Longest ORF"
		/codon_start=1
		/product="GH20904p"
		/protein_id="AAL28306.1"
		/db_xref="GI:16768174"
Oy	1 GACCACGTSCGSGCATG 18    :::  ::        :	/db_xref="FBgn0033869" /translation="MMSFKSFWLSTLSIAVAAYAOIAPGSNAVLPPTKNGDYSEE PKTFPFGPPGRPAARCPRPAPPGTFRNIPGGQGDHVAHPGNPIFYAVDPEPRFA NDAAHKRSSDGVDVTGEVRVMEDRTOIVRYADWKTYHADVSYEATNYFPGLPOP GARGGGGAGAGAGGYKYX"
Dn	454 GTCCACGTGCCGCATG 471  	
RESULT_15		
LOCUS	AR480461/c	1008 bp DNA linear PAT 14-MAY-2004
DEFINITION	Sequence 1365 from patent US 6699703.	
ACCESSION	AR480461	
VERSION	AR480461.1 GI:47239423	
KEYWORDS	.	
SOURCE	Unknown.	
ORGANISM	Unclassified. 1 (bases 1 to 1008) Doucette-Stamm,L., Bush,D., Zeng,Q., Opperman,T. and Housewart,C.E. Nucleic acid and amino acid sequences relating to Streptococcus pneumoniae for diagnostics and therapeutics Patent: US 6699703-A 1365 02-MAR-2004; location/Qualifiers 1..1008 /organism="unknown" /mol_type="genomic DNA"	
JOURNAL FEATURES		
TITLE		
AUTHORS		
REFERENCE MATCHES		
Query Match	Best Local Similarity 86.7%; Score 15.6; DB 6; Length 1008;	
Matches	15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;	
Ox	1 GACCACGTSCGSGCATG 18      :: :: :      :	
Dn	556 GACCACGTGCCGCATG 539  	
RESULT_16		
LOCUS	AX567274/c	1188 bp DNA linear PAT 29-NOV-2002
DEFINITION	Sequence 473 from Patent WO02077021.	
ACCESSION	AX567274	
VERSION	AX567274.1 GI:26001699	
KEYWORDS	. Streptococcus pneumoniae Streptococcus pneumoniae Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.	
SOURCE		
ORGANISM		
REFERENCE	1 Maegmani,V., Tettelin,H. and Fraser,C. Streptococcus pneumoniae proteins and nucleic acids Patent: WO 02077021-A 473 03-OCT-2002; Chiron Spa (IT) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)	
FEATURES	Location/Qualifiers 1..1188	
SOURCE		



/organism="Streptococcus pneumoniae"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:1313"

ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 1188;  
Best Local Similarity 83.3%; Pred. No. 7.9e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACGTSCCGGCATG 18  
|||||:|||||  
Db 739 GACCACGTGCCGCATG 722

RESULT 17

E30250 1408 bp DNA linear PAT 18-JUN-2001  
LOCUS Process for producing biotin.  
DEFINITION E30250  
ACCESSION E30250.1 GI:13021423  
VERSION JP 199276175-A/1.  
KEYWORDS Sphingomonas sp.  
SOURCE Sphingomonas sp.  
ORGANISM Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; Sphingomonas.

REFERENCE 1 (bases 1 to 1408)  
AUTHORS Tomoyasu, K., Fujio, S. and Masatoshi, S.  
TITLE Process for producing biotin  
JOURNALS Patent: JP 199276175-A 1 12-OCT-1999;  
SUMITOMO CHEM CO LTD  
OS Sphingomonas sp.  
PN JP 199276175-A/1  
PD 12-OCT-1999  
PF 31-MAR-1998 JP 1998086972

PR C12N15/09, C12N1/21, C12P17/18// (C12N15/09, C12R1:01), C12P17/18, PC  
C12R1:19),  
PC C12N15/00, (C12N15/00, C12R1:01)  
CC Strandedness: Double;  
CC Topology: Linear;  
FH Key Location/Qualifiers  
FT CDS 202..1365.  
Location/Qualifiers  
1..1408  
/organism="Sphingomonas sp."  
/mol\_type="genomic DNA"  
/db\_xref="taxon:28214"

ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 1408;  
Best Local Similarity 83.3%; Pred. No. 7.7e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACGTSCCGGCATG 18  
|||||:|||||  
Db 1057 GACCTCGTCCGCGCATG 1074

RESULT 18

E30254 1408 bp DNA linear PAT 18-JUN-2001  
LOCUS Process for producing biotin.  
DEFINITION E30254  
ACCESSION E30254.1 GI:13021427  
VERSION JP 199276175-A/5.  
KEYWORDS Sphingomonas sp.  
SOURCE Sphingomonas sp.  
ORGANISM Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; Sphingomonas.

REFERENCE 1 (bases 1 to 1408)  
AUTHORS Tomoyasu, K., Fujio, S. and Masatoshi, S.  
TITLE Process for producing biotin

JOURNAL Patent: JP 1999276175-A 5 12-OCT-1999;  
SUMITOMO CHEM CO LTD  
OS Sphingomonas sp.  
PN JP 199276175-A/5  
PD 12-OCT-1999  
PF 31-MAR-1998 JP 1998086972

PR C12N15/09, C12N1/21, C12P17/18// (C12N15/09, C12R1:01), C12P17/18, PC  
C12R1:19),  
PC C12N15/00, (C12N15/00, C12R1:01)  
CC Strandedness: Double;  
CC Topology: Linear;  
FH Key Location/Qualifiers  
FT CDS 202..1365.  
Location/Qualifiers  
1..1408  
/organism="Sphingomonas sp."  
/mol\_type="genomic DNA"  
/db\_xref="taxon:28214"

FEATURES  
source

ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 1408;  
Best Local Similarity 83.3%; Pred. No. 7.7e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACGTSCCGGCATG 18  
|||||:|||||  
Db 1057 GACCTCGTCCGCGCATG 1074

RESULT 19

AR215256 1408 bp DNA linear PAT 25-SEP-2002  
LOCUS Sequence 5 from patent US 6410293.  
DEFINITION AR215256  
ACCESSION AR215256.1 GI:23313427  
VERSION AR215256.1  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1408)  
AUTHORS Mukumoto, F., Nishio, S., Akimaru, J. and Mitsuada, S.  
TITLE DNA fragments containing biotin biosynthease gene and use of the same  
JOURNALS Patent: US 6410293-A 5 25-JUN-2002;  
Location/Qualifiers  
1..1408  
/organism="unknown"  
/mol\_type="genomic DNA"

FEATURES  
source

ORIGIN

Query Match 86.7%; Score 15.6; DB 6; Length 1408;  
Best Local Similarity 83.3%; Pred. No. 7.7e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GACCACGTSCCGGCATG 18  
|||||:|||||  
Db 1057 GACCTCGTCCGCGCATG 1074

RESULT 20

AR215257 1408 bp DNA linear PAT 25-SEP-2002  
LOCUS Sequence 7 from patent US 6410293.  
DEFINITION AR215257  
ACCESSION AR215257.1 GI:23313428  
VERSION AR215257.1  
KEYWORDS Unknown.  
SOURCE Unknown.  
ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1408)  
AUTHORS Mukumoto, F., Nishio, S., Akimaru, J. and Mitsuada, S.

REFERENCE	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.									
AUTHORS	1 Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M., Katsagiri,F., Qian,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.									
TITLE	Plant genes involved in defense against pathogens									
JOURNAL	Patent: WO 03000898-A 6310 03-JAN-2003; Syngenta Participations AG (CH)									
FEATURES	Location/Qualifiers									
source	1..2000 /organism="Oryza sativa" /mol_type="unassigned DNA" /db_xref="taxon:4530"									
ORIGIN										
Query Match	86.7%;	Score 15.6;	DB 6;	Length 2000;						
Best Local Similarity	83.3%;	Pred. No. 7.2e+03;								
Matches	15;	Conservative	2;	Mismatches	1;	Indels	0;	Gaps	0;	
OY	1 GACCACGCTCCGGCGCATG 18      :									
Db	1089 GGCCACGTCGCCGCGCATG 1106									
RESULT 23										
LOCUS	BD139028 2048 bp DNA linear PAT 18-SEP-2002									
DEFINITION	Viruslike-associated nucleic acid sequences and uses thereof.									
ACCESSION	BD139028									
VERSION	BD139028.1 GI:23233973									
KEYWORDS	JP 2002505849-A/76.									
SOURCE	Pseudomonas aeruginosa									
ORGANISM	Pseudomonas aeruginosa Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.									
REFERENCE	1 (bases 1 to 2048)									
AUTHORS	Ausubel,F., Goodman,H.M., Ratme,L.G., Miklos,S.M., Tan,M.W., Cao,H., Drenth,E. and Isongalis,U.									
TITLE	Viruslike-associated nucleic acid sequences and uses thereof									
JOURNAL	Patent: JP 2002505849-A 76 26-FEB-2002; THE GENERAL HOSPITAL CORP									
COMMENT	OS Pseudomonas aeruginosa PN JP 2002505849-A/76 PD 26-FEB-2002 PF 25-NOV-1998 JP 2000522270 PR 25-NOV-1997 US 60/066517 PI FREDERICK AUSUBEL, HOWARD M GOODMAN, LAURENCE G RAHME PI									
FEATURES	SHALINA MARHJAN MIKLOS, PI MAN WAH TAN, HUI CAO, ELIANA DREHKARD, JOHN TSONGALIS PC C12N15/09,A61K45/00,A61P31/04,C07K14/21,C12Q1/02,C12Q1/68, PC G01N33/15, PC G01N33/50//C12P21/02,(C12N15/09,C12R1.385),C12N15/00,(C12N15/00,C12R1.385) PC 00,C12R1.385) CC Viruslike-associated nucleic acid sequences and uses thereof FH key Location/Qualifiers FT source 1..2048 FT /organism='Pseudomonas aeruginosa'.									
ORIGIN	1..2048 /organism="Pseudomonas aeruginosa" /mol_type="genomic DNA" /db_xref="taxon:287"									
Query Match	86.7%;	Score 15.6;	DB 6;	Length 2048;						
Best Local Similarity	83.3%;	Pred. No. 7.1e+03;								
Matches	15;	Conservative	2;	Mismatches	1;	Indels	0;	Gaps	0;	
OY	1 GACCACGTCGCCGCGCATG 18      :									
Db	1069 GACCACGTCGCCGCGCATG 1086									

RESULT 24  
LOCUS AR198851 2048 bp DNA linear PAT 20-APR-2002  
DEFINITION Sequence 136 from patent US 6355411.  
ACCESSION AR198851  
VERSION AR198851.1 GI:20248925  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
FEATURES  
REFERENCE 1 (bases 1 to 2048)  
AUTHORS Ausubel,F., Goodman,H.M., Rahme,L.G., Mahajan-Miklos,S., Tan,M.-W.,  
Cao,H., Drenth,E. and Tsongalis,J.  
TITLE Virulence-associated nucleic acid sequences and uses thereof  
JOURNAL Patent: US 6355411-A 136 12-MAR-2002;  
FEATURES  
source  
1..2048  
Location/Qualifiers  
/organism="unknown"  
/mol\_type="unassigned DNA"  
ORIGIN  
Query Match 86.7%; Score 15.6; DB 6; Length 2048;  
Best Local Similarity 83.3%; Pred. No. 7.1e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GACCACGTSCCGGCATG 18  
|||||:|||||  
Db 1069 GACCACGTSCCGGCATG 1086  
RESULT 25  
LOCUS CQ789114/c 2395 bp DNA linear PAT 29-MAR-2004  
DEFINITION Sequence 205 from Patent EP1400592.  
ACCESSION CQ789114  
VERSION CQ789114.1 GI:45822682  
KEYWORDS  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1  
AUTHORS Kunesh,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Barash,S.C.,  
Fannon,M. and Dougherty,B.A.  
TITLE Streptococcus pneumoniae polynucleotides and sequences  
JOURNAL Patent: EP 1400592-A 205 24-MAR-2004;  
FEATURES  
source  
1..2395  
Location/Qualifiers  
/organism="unidentified"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"  
ORIGIN  
Query Match 86.7%; Score 15.6; DB 6; Length 2395;  
Best Local Similarity 83.3%; Pred. No. 6.9e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GACCACGTSCCGGCATG 18  
|||||:|||||  
Db 1257 GACCACGTSCCGGCATG 1240  
RESULT 26  
LOCUS AR218973/c 2395 bp DNA linear PAT 25-SEP-2002  
DEFINITION Sequence 205 from patent US 6420135.  
ACCESSION AR218973  
VERSION AR218973.1 GI:23319907  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
FEATURES  
REFERENCE 1 (bases 1 to 2395)

AUTHORS Kunesh,C.A., Choi,G.H., Dillon,P.S., Rosen,C.A., Barash,S.C.,  
Fannon,M.R. and Dougherty,B.A.  
TITLE Streptococcus pneumoniae polynucleotides and sequences  
JOURNAL Patent: US 6420135-A 205 16-JUL-2002;  
FEATURES  
source  
1..2395  
Location/Qualifiers  
/organism="unknown"  
/mol\_type="genomic DNA"  
ORIGIN  
Query Match 86.7%; Score 15.6; DB 6; Length 2395;  
Best Local Similarity 83.3%; Pred. No. 6.9e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GACCACGTSCCGGCATG 18  
|||||:|||||  
Db 1257 GACCACGTSCCGGCATG 1240  
RESULT 27  
LOCUS BD003885/c 2395 bp DNA linear PAT 31-JAN-2002  
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.  
ACCESSION BD003885  
VERSION BD003885.1 GI:18631846  
KEYWORDS  
SOURCE unidentified  
ORGANISM unidentified  
REFERENCE 1 (bases 1 to 2395)  
AUTHORS Kunesh,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Barash,S.C.,  
Fannon,M. and Dougherty,B.A.  
TITLE Polynucleotide of Streptococcus pneumoniae and sequence  
JOURNAL Patent: JP 2001501833-A 205 13-FEB-2001;  
COMMENT HUMAN GENOME SCIENCES INC  
OS Unidentified  
PN JP 2001501833-A/205  
PD 13-FEB-2001  
PF 30-OCT-1997 JP 1998520718  
PR 31-OCT-1996 US 60/029960  
PI CHARLES A KUNESH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI  
STEVEN C BARASH,  
MICHAEL FANNON, BRIAN A DOUGHERTY  
PC C12N15/09, A01K67/027, C07K14/315, C07K16/12, C12N1/15, C12N1/19,  
PC C12N1/21,  
PC C12N5/10, C12P21/02, C12Q1/68, G06F17/30, C12N15/00, C12N5/00, PC  
G06F15/40  
CC Strandedness: Double;  
CC Topology: Linear;  
FH Key  
FT source  
1..2395  
Location/Qualifiers  
/organism="unidentified".  
ORIGIN  
Query Match 86.7%; Score 15.6; DB 6; Length 2395;  
Best Local Similarity 83.3%; Pred. No. 6.9e+03;  
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GACCACGTSCCGGCATG 18  
|||||:|||||  
Db 1257 GACCACGTSCCGGCATG 1240  
RESULT 28  
LOCUS HSH12B 3233 bp DNA linear PRI 20-MAY-1996  
DEFINITION Homo sapiens H12B gene.  
ACCESSION X81326

```

VERSION      X81326.1 GI:587211
KEYWORDS     repeated element.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE    1
AUTHORS     Ryakov,A.P., Prosyak,M.I., Kupriyanova,N.S., Netchvolodov,K.K. and
              Limborska,S.A.
TITLE        A new multi-locus DNA fingerprinting probe: K25
JOURNAL      DNA Res. 2 (3), 151-152 (1995)
MEDLINE      96038259
PUBMED       8581743
REFERENCE    2 (bases 1 to 3233)
AUTHORS     Prosyak,M.I.
TITLE        Direct Submision
JOURNAL      Submitted (03-SEP-1994) M.I. Prosyak, Inst of Molecular Genetics,
              Russian Academy of Sciences, Kurchatov sq, 123182 Moscow, RUSSIA

FEATURES     source
              Location/Qualifiers
                1..3233
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /clone="H12B"
                /haploidy="diploid"
                /cell_type="lymphocytes"
                /clone_lib="human genomic DNA in lambda EMBL-3 phage"
                /dev_stage="adult"
                38..540
                repeat_region
                /evidence=experimental
                576..5840
                repeat_region
                /evidence=experimental

ORIGIN
Query Match      86.7%; Score 15.6; DB 9; Length 3233;
Best Local Similarity 83.3%; Pred. No. 6.6e+03;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 GACCAAGTSCCGGCATG 18
      |||||:|:|:|:|:|:|
Db      829 GACCACTGCCCGCATG 846

RESULT 29
AF227905      4927 bp mRNA linear PRI 30-APR-2001
DEFINITION   Homo sapiens UDP-glucose:glycoprotein glucosyltransferase 1
ACCESSION    AF227905
VERSION      AF227905.1 GI:7670745
KEYWORDS
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 4927)
AUTHORS     Arnold,S.M., Fessler,L.I., Fessler,J.H. and Kaufman,R.J.
TITLE        Two homologues encoding human UDP-glucose:glycoprotein
              glucosyltransferase differ in mRNA expression and enzymatic
              activity
JOURNAL      Biochemistry 39 (9), 2149-2163 (2000)
MEDLINE      20160491
PUBMED       10694380
AUTHORS     Arnold,S.M.
TITLE        Direct Submision
JOURNAL      Submitted (22-JAN-2000) Biological Chemistry and Howard Hughes
              Medical Institute, University of Michigan, 1150 West Medical Center
              Drive, Ann Arbor, MI 48109-0650, USA

FEATURES     source
              Location/Qualifiers
                1..4927
                /organism="Homo sapiens"
                /mol_type="mRNA"

```

```

CDS
/db_xref="taxon:9606"
/issue_type="liver"
/dev_stage="fetal"
39..4706
/feature="HUGT1"
/codon_start=1
/product="UDP-glucose:glycoprotein glucosyltransferase 1
precursor"
protein_id="AAE66232.1"
/db_xref="GI:7670746"
/translation="MGCKDASGACAGALPTVGVCYKMGVLVVLPTVLELSSSVKADS
KAITTSYTKMSTPLLEASERFLAEDSOEFNNFVSAQNSISDHGIDYGYHAI
LEAFORSLPLOONLEKFCILSRYSATIOAFQOIADEPGECSNFFVHGKTCB
SDTLERALLTASERPKPLFKGHRVSSNPSBPVYFEISGESEENSRHOLISKS
NAGKINVFPHYIFNPKBEFVLSGVEELASTERYKADDDQVKTETVNTYIGN
DPIDVEQFLFGKRLDHPLEGOQLKRLRLVSEINEMAPLKWQLODSFQPAKI
LASFEVALVVMKDSQNFPEPKARAIKTVASSELREVEENQKFFEGTGLPGDSA
LEFINGLMDQDQDIFSLPVLNREARVMEGLRLGIEGLINVLKLANIQPSADVA
VDLRSPALISVMNLLEVDGRYNSNPSLSQELLRPFPGVIRQIKNLHNVFVIDPAE
TTBALNANTAEMLFSNHLPLKIGFI FVYNDESDVDGMDAGVAVIRAYVVAQVDVH
AFQTLTHIYNKVRGKRVKHVAVSLEKTPYVEVNSIIGISAYDRNRKARGYTE
QTVGRLPVVLFNMPERBOLDPELEITMHLILETTFQRAVYLGLPHQDVV
EYIMNOPVVRINRSILITAEADYLDLTASNFEVDYARFTILDSGCKTAAVANSN
YLTKGMSKRIYDPSFIRPVTWIVGDPSGROLVDAIKHOKSNVRSIMNN
PAKISYENROI SPATWALOTOTSNAKMPITKMEGABALAAAGADIAPFVGGM
DPSLFKEFSSKNDFTLSAAVTCRDVLKAKGORAVISNGRIIGPEDESELNODP
HLENIILKTSQKIKSHIQQLRVEEDVADLVKVALISAQKGDPRLEYOFER
HSAIKLRPKGETFYDVAADVPTVTRAQRLAPLVLVLAQLINNNLPEVNNQSKSD
MPLKSPRYVLEPEISFTSDNSFAKGP IAKFLDMPQSPPLFTLNTPESSWVESVRP
YDLDNIYLEVDVVAAYELEYLLGHCYDITGGPRLGTLGTSANPVVDIT
VMANUGYFOALKANGAWILRLRKGRSDIRIYSHDGTDPADDEVIVLNNKSKI
IKAVQKKADVWNEOLSDGTSEBSEFMSFKGPTGQTEBEVKQKDDIINIYSVA
SGHLYERFLRLMILSVLAKNTKTPVKEFWFLNKYLSPTKERTPYMANTNFOYELVQK
WPRWILHOOTEKORLIWGYKILFLDLVLPVLVDKFLPYDADQIVTDLKEADENLDA
PYGATPFCDSREMDGYRFWSGYWASHLAGRKXIKALVYVDLAKRKRIKAAAGBRLNG
OYGLSDQDPSLSINLDDLPNNMHQVPIKSLPQEMLCETWCDDASKKAKITIDLN
NPKMKERKLEAAVIVPEWQDYDEINQILQIRPKKXETGALYEXKTEPSREBPQKR
BEL"
39..164
165..4703
/product="UDP-glucose:glycoprotein glucosyltransferase 1"

ORIGIN
Query Match      86.7%; Score 15.6; DB 9; Length 4927;
Best Local Similarity 83.3%; Pred. No. 6.1e+03;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 GACCAAGTSCCGGCATG 18
      |||||:|:|:|:|:|:|
Db      24 GACCAAGGCGCGGCATG 41

RESULT 30
AX405692      5042 bp DNA linear PAT 14-JUN-2002
DEFINITION   Sequence 107 from Patent WO0222660.
ACCESSION    AX405692
VERSION      AX405692.1 GI:21438823
KEYWORDS
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1
AUTHORS     Tang,Y.T., Liu,C., Zhou,P., Aundil,V., Zhang,J., Zhao,Q.A., Ren,F.,
              Xue,A.J., Yang,Y., Wehman,T. and Dermanac,R.T.
TITLE        Novel nucleic acids and polypeptides
JOURNAL      Patent: WO 0222660-A 107 21-MAR-2002;
              HYSEQ, INC. (US)

FEATURES     source
              Location/Qualifiers
                1..5042
                /organism="Homo sapiens"
                /mol_type="unassigned DNA"

```

CDS

```

/db_xref="taxon:9606"
164..4831
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAD34799.1"
/db_xref="GI:21438824"
/translation="MGCKGDASGACAAGALPTMGVCYKMGVLVVLTVLMLFSSVKADS
KATITSLITTKMFSPTLLLEASERFLAEDSQEFNWFVASEQNISSSDHGTDSYTHAI
LEAFOPLSPLOQNLPKFCCLRSYSATIAFOQIADDEPPGCSNFFSVHGKTCSE
SDTEBALLTASERPKPLFKGDHRYSPSNDESPVIFYSEIGSEEFNSFROLISKS
NAGKINVPFRHYIENPKRKYVLSGVGVELAIKSTYKADDPQVGTETVNTVIGEN
DPIDVEOGFIPKIRDLHPDLBEGQLKELRKLHLYKSTENMAPLKYMQLODSFQTAARI
LASPEVALVVMKDLSONPPTKAAITKTAVSSELTREVENOKTFEKGTLGLQPADSA
LFTNGLMDDLTDIDFSLPDLVLRNEARMEGLHRLGELSHNVLYKLNQPSADYA
VDLRSPAISVNNLEVDSTRYNSWSSLOELLRPTFGVIRQIRKLNLMVFIYDPAHE
TTAELMNTAMFPLSNHILRIGLFFVNDSEDDVGMQDAGVAVRAYVVAOEVDDYH
APOTLPHIYKVRGTGEKVHVSUVEKKYPVVEVNSITIGIDSAVDNRKREAGYVE
QTGGLPVLVLFNGMPFERQLDDELTTMKILETTTFFQRAVYLGELPHQDAV
EYIMNQNVNVRINSRILTAERDYLDTASNNFFVDDYARFTLLDSQKTAAVANSKN
YLTCKGMSKEIYDDSFIRPVTFWIVGDFPSGROQLYDAIKHOKSSNNVRI SMIN
PAKEISYENTQISRAIWAALQOTSNNAKNFTMAKEGAABALAGADIAEFSVGM
DPSLPEKVEFSSKMDFILSHAVCRDVIKIKGORAVISNGRIIGLEDESELFNODP
HLENTILKTSQKIKSHIQQLRVEEDVASDVMKVYALLSAQKEDPRIEYQFEER
HSAIKLRKEGETFDVAVVDPVTEARQLAPLLVLVLAQLINNNLRVFNKCSKLSB
MPLKSPRYVLEBEISTSDNSPAKPIAKFLDMPQSPPLTLNLTPESNMVSVRTP
YDLNITYLEEVDSVAAEYLEYLLLEGHCYDITTCQPPRGLQFTLGTSANPVIYDTI
VMANIGFOLKANGAWILRLRKGSEDIYRIYSHDGTSPDPADEVIVIANPKSKI
IKYVOKKADMANVEDILSDGTSENEGSPKSGFTGOKTEEVKDDKDDIINISVA
SGHLIERFLIMLSVKATKTPTPKYKFFLKQYLSPTPKETIPYMANEYNFQYELVQYK
WPRMLHQOTEKQRIWGYKILFLDVLEPLVVDKFLFVDAQIVRTDLKEILDVFLDGA
PYGVTPEQDSRREMDGYRFWKSQYMAHLAGRKYHISALYVDLKKFRKIAGDBRLG
QYQGLSQDPPNSLSNLDQDLNNMTHQVPIKSLPOEMLMCEMTCDDASKKRAKTIIDLGN
NPMTKEPKLEAAVRIPEMQDYDEIKQLQIRFOKEKETGALYKEXTKEPSRSGPQKR
BEL"

```

ORIGIN

```

Query Match      86.7%; Score 15.6; DB 6; Length 5042;
Best Local Similarity 83.3%; Pred. No. 6e+03;
Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 GACCACGTCGCGGCATG 18
      |||||:|||||
      149 GACCACGCGCGGCATG 166

```

Search completed: July 20, 2005, 16:45:08  
 Job time : 903.429 secs

**THIS PAGE IS BLANK**